Emphasizing Group Project-Based Learning Elements in Designing Instructional Material for Univariate Time Series Analysis and Forecasting

Mahayaudin M. Mansor

School of Mathematical Sciences
College of Computing
Informatics and Mathematics
Universiti Teknologi MARA
40450 Shah Alam, Malaysia

Email: maha@uitm.edu.my

Abstract

Project Based Learning (PBL) is a systematic teaching method in which students learn by engaging in real-world problems and applying the essential time series knowledge and univariate modeling concepts learned to solve the problems. In addition to PBL, this contribution emphasizes the importance of working together in groups so that students can improve their social skills, critical thinking, and numeracy skills. This is important because analysts and statisticians generally enjoy working in silos. In addition, univariate time series modeling and model validation procedures can be complicated for many students. Hence, the proposed instrumental material is carefully developed, comprising comprehensive step-by-step tasks related to univariate time series analysis using RStudio. The proposed hands-on group assignment outlines instructions and 13 steps centered on the seven Group PBL (GPBL) elements of planning, fieldwork, implementation, presentation, assessment-reflection, collaboration, and effective communication. Planning involves discussing data searches, setting targets and timelines, and distributing tasks among team members. While fieldwork in GPBL refers to gathering and compiling data and resources, its implementation requires team members to execute the outlined tasks meticulously to avoid mistakes. Students are required to present their group work to an audience, and peers will provide formative feedback with instructor guidance for group reflection and improvement. The other advantages are the range of R programming commands for univariate times series modeling and forecasting, and the qualitative and quantitative diagnostic criteria for residual assumptions are given in Steps 1 to 13. To the best of my knowledge, no similar instrumental material has been produced. Besides patenting, the benefits of using the proposed instrumental material for students include more profound knowledge about the topic, higher-order thinking skills, collaborative learning skills, and real-life applications that have value for public products, including publication.

Keywords: Real-world application, student engagement, social skills, critical thinking, systematic teaching method.

TEACHING AND LEARNING POSTER IDEAS
TALPI 2023
"TOWARDS INNOVATIVE GLOBAL TRENDS IN EDUCATION "
VIRTUAL COMPETITION

EMPHASIZING GROUP PROJECT BASED LEARNING ELEMENTS IN DESIGNING INSTRUCTIONAL MATERIAL FOR UNIVARIATE TIME SERIES ANALYSIS AND FORECASTING

Mahayaudin M. Mansor1

[1] *School of Mathematical Sciences, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, 40450 Shah Alam, Malaysia*
Corresponding email: maha@uitm.edu.my

Project Based Learning (PBL) is a systematic teaching method in which students learn by engaging in real-world problems and applying the essential time series knowledge and univariate modelling concepts learned to solve the problems. In addition to PBL, this contribution emphasizes the importance of working together in groups so that students can also improve their social skills in addition to critical thinking and numeracy skills. This is important because analysts and statisticians, in general, enjoy working in silos. In addition, univariate time series modelling and model validation procedures can be complicated for many students. Hence, the proposed instrumental material is carefully developed comprising comprehensive step-by-step tasks related to univariate time series analysis using *R*Studio. The proposed hands- on group assignment outlines a list of instructions and 13 steps centered on the seven Group PBL (GPBL) elements of planning, fieldwork, implementation, presentation, assessment-reflection, collaboration, and effective communication. Planning involves discussing data searches, setting targets and timelines, and distributing tasks among team members. While fieldwork in GPBL refers to gathering and compiling data and resources, its implementation requires team members to execute the outlined tasks meticulously to avoid mistakes. Students are required to present their group work to an audience and formative feedback will be provided by peers with instructor guidance for group reflection and improvement. The other advantages are the range of *R* programming commands for univariate times series modelling and forecasting, and the qualitative and quantitative diagnostic criteria for residual assumptions are given in Steps 1 to 13. To the best of my knowledge, there is no similar instrumental material that has been produced. Besides patenting, the benefits of using the proposed instrumental material for students include deeper knowledge about the topic, higher-order thinking skills, collaborative learning skills, and real-life applications that have value for public products including publication.

Keywords: Real-world application, student engagement, social skills, critical thinking, systematic teaching method.

UNIVERSITI TEKNOLOGI MARA

Group Assignment:
Univariate Time Series Modelling Techniques of the Course Time Series Analysis and Forecasting

Instructions:

1. **Perform this assignment in a group of three (3) members consisting of at least one male and one female, if possible. Assign at least 3 tasks to one (1) team member.**
2. **Follow all the steps and answer all questions in English.**
3. **Prepare a plan on how to execute this group work together with a timeline and description of responsibilities for each member and as a team.**
4. **Present the plan and the group work during the 15-minute group presentation.**
5. **Submit the final report in week 10 by incorporating the feedback received from the instructor and audience during the presentation.**
6. **For each step and question, you must provide quantitative and qualitative evidence (e.g. statistical tests and charts) in your report and interpret the results appropriately to earn good marks. Request the rubric to manage your expectations for the final report.**
7. **You are encouraged to use tables and figures to present your answers.**
8. **The literature must be properly cited. Use APA style for the literature cited in this work.**
9. **The execution of this work must be done using *R* computing only.**
10. **The time series data used must represent a genuine real-world problem.**

Step 1: Data search, problem statement, and literature

In a group, brainstorm and decide about the selection of a time series to be used for this study. Download the time series from any reliable and free publicly available data sources or official websites. Note that the chosen time series must be unique from other groups. The time intervals for the time series can be in second, minute, hour, daily, monthly, quarterly, or yearly. The length of the time series must be at least from the year 1999 or 2000. Save the data file in CSV or Excel format before importing the data for analysis. Prepare to submit a copy of the data file while submitting the final report. Also, report the justification for choosing the series, and clearly discuss the problem statement following citations of at least 10 literature beginning in 2000 related to the chosen time series.

[10/100 marks]

Step 2: Importing the time series to *RStudio*

Once the time series is uploaded, you may use '*time_series*' as its caller name. Use '*ts()*' function to define the data as time series objects. Prepare the *R* file to be submitted together with the final report.

[2/100 marks]

Step 3: Missing values

Report for any missing values. If there is any, impute the missing values and note the treatment used with at least one citation in the report. Use *summary(time_series)*.

[2/100 marks]

Step 4: Summary statistics

Report the summary statistics in a table including the unit measurement, length, minimum, mean, median, maximum, standard deviation, and skewness. Also, use *head(time_series)* and *tail(time_series)* to display the first and last six observations of the time series.

[10/100 marks]

Step 5: Shape of distribution and outlier

Use *hist(time_series)* and *boxplot(time_series)* to plot a histogram and a boxplot for the time series. Based on the histogram and boxplot, explain the skewness of the distribution in relation to normal distribution. State any suitable treatment for the skewed data in relation to the forthcoming analysis, if needed. Next, report any outlier using a built-in function in *R*, *boxplot.stats(time_series)$out* for the series. In the report, explain the event(s) associated with the outlier(s) supported by at least one literature. Although your time series is not skewed or contains no outliers, please note the conditions in the report.

[6/100 marks]

Step 6: Time series plot and its components

Make sure that the time interval of the series is regular or uniform. Otherwise, interpolate the series. Next, produce the following plots to assist with the description of time series components: *plot.ts(time_series)*; *acf(time_series) or acf(c(time_series))*; and *ggseasonplot(time_series)*. Based on the three (3) plots, completely describe all the time series components for the series.

[7/100 marks]

Step 7: Fit suitable time series models

Fit one (1) model of the following models from Chapter 2 to the chosen time series. Use all points in the series.

(a) One of the Naïve models,
(b) One of the methods of averages, and
(c) One of the exponential smoothing techniques.

State the general equation used in the report accordingly. Justify the selections by commenting on the suitability of the chosen models for the time series. Cite at least one past study that used the chosen models.

[9/100 marks]

Step 8: The estimated coefficient(s)

Report the estimated model coefficient(s) for each model. Report the fitted models accordingly.

[3/100 marks]

Step 9: How well do the models fit the data?

Plot the fitted values (e.g. *model$fitted.values*) and observed values in a single time series plot. Use different colours for the two lines (e.g. blue and black). Observe and comment on the gap between the two lines or the discrepancies between the fitted and the observed values for all fitted models.

[6/100 marks]

Step 10: Error measures

Extract the residuals to a caller name (e.g. *residuals_model=model$residuals*) from each model. Use the residuals to estimate error measures of ME, RMSE, MAE, MPE, MAPE, and MASE. Use *accuracy()* function for the in-sample or re-substitution error. Report the results in a table. For each error measure, bold the smallest value and conclude which one out of the three fitted models has the most smallest error measures. Relate the explanation of estimates with the plots in Step 9.

[9/100 marks]

Step 11: Error assumptions and diagnostics

Based on Steps 9 and 10, choose one fitted model that has the closest estimated values to the observed. Use the residuals from the chosen model to produce the following plots and statistical tests for the error assumptions.

(a) Normality [*qqnorm()* with *qqline* and Anderson-Darling test or any other normality tests e.g. Kolmogorov-Smirnov, Shapiro-Wilk, Jarque-Bera]
(b) Independently distributed (no discernible pattern) [sequence plot and Ljung-Box test]
(c) Homoscedasticity (constant variance) [scatterplot between *standardized residuals versus fitted values and Breusch-Pagan test]
(d) Absence of lag dependence (serially independent) [ACF plot and Durbin-Watson test or Breush-Godfrey test]
(e) Absence of volatility (additional for financial time series) [plot of *acf*(*residuals_model*^2)].

Report all the qualitative and quantitative assessments in tables. Comment/interpret each of the plots and *p*-value from the statistical tests in relation to the modelling assumptions. It is compulsory to cite at least one literature for each assumption. **standardized residuals= residuals_model/sd(residuals_model)*

[20/100 marks]

Step 12: The best model

Based on Steps 9, 10, and 11 state the best model. Clearly explain why the stated model is the best to represent the chosen time series. Cite at least 3 other past studies that used the same best model for the chosen time series.

[6/100 marks]

Step 13: Forecasting, future trends, and recommendation

Finally, use the best model in Step 12 to estimate at least 10-step ahead for the time series. Show the ex-ante (out-of-sample) forecast values in a table and produce a line plot for the future values alongside the observed and the fitted values from the best model in a single time series plot. Please note that the table (or the chart) reports all the forecast values including the dates and their confidence intervals of say 80% and 95% obtained from e.g. *forecast()* function. Comment about future trends and propose a solution or recommendation for the problem. Cite at least 2 past studies to support the solution or recommendation.

[10/100 marks]

Pejabat Perpustakaan
Library Office

Universiti Teknologi MARA
Cawangan Perak
Kampus Seri Iskandar
32610 Bandar Baru Seri Iskandar,
Perak Darul Ridzuan, MALAYSIA
Tel: (+605) 374 2093/2453 Faks: (+605) 374 2299

UNIVERSITI TEKNOLOGI MARA

Surat kami : 700-KPK (PRP.UP.1/20/1)
Tarikh : 20 Januari 2023

Prof. Madya Dr. Nur Hisham Ibrahim
Rektor
Universiti Teknologi MARA
Cawangan Perak

Tuan,

PERMOHONAN KELULUSAN MEMUAT NAIK PENERBITAN UiTM CAWANGAN PERAK MELALUI REPOSITORI INSTITUSI UiTM (IR)

Perkara di atas adalah dirujuk.

2. Adalah dimaklumkan bahawa pihak kami ingin memohon kelulusan tuan untuk mengimbas (*digitize*) dan memuat naik semua jenis penerbitan di bawah UiTM Cawangan Perak melalui Repositori Institusi UiTM, PTAR.

3. Tujuan permohonan ini adalah bagi membolehkan akses yang lebih meluas oleh pengguna perpustakaan terhadap semua maklumat yang terkandung di dalam penerbitan melalui laman Web PTAR UiTM Cawangan Perak.

Kelulusan daripada pihak tuan dalam perkara ini amat dihargai.

Sekian, terima kasih.

"BERKHIDMAT UNTUK NEGARA"

Setuju.

27.1.2023

PROF. MADYA DR. NUR HISHAM IBRAHIM
REKTOR
UNIVERSITI TEKNOLOGI MARA
CAWANGAN PERAK
KAMPUS SERI ISKANDAR

Saya yang menjalankan amanah,

SITI BASRIYAH SHAIK BAHARUDIN
Timbalan Ketua Pustakawan

nar