In this installment 2 you have three tasks: (1) Fix or add anything I asked you to fix or add to the part done in installment 1. (2) Keep editing the paper according to the instructions given below, and then (3) do the new work described below. Then upload according to instructions

Please, notice that starting with section IV of the paper, you will work only with n-12 observations. That is, if your data set has 240 observations, do the data transformations, ACF, PACF just with the first 228. The last 12 will be used to compare your forecasts with the actual data, to see how well your methods forecast.

(1) Fix or add anything I asked you to fix or add to the part done in installment 1. (2) Keep editing the paper according to the instructions given below

(1) Start organizing the paper to put it in its final form (creating placeholders for future sectinos). You will be adding things to it as the quarter advances. Here are the parts that your paper will have. Write all the Section titles and numbers that are not there yet.

- A front page with your title, your name, the class name and your student ID. Nothing else in it. (should have been done in installment 1)
- On the second page, start Section 1. Introduction. In this section you state the objective of the paper. Then you say what data set you chose to satisfy the objectives of the project. Your hypothesis. Why did you make that hypothesis. What leads you to believe it is true? References must be marked with either a number [1][2] to find them easily in the reference section at the end of the paper, or with (Author's last name, year). Write nicely and with style. All this should have been done in installment 1. In this new installment, **ADD to your introduction now, the purpose of this paper: to compare several time series models to see which one gives the best forecasting performance in the short and the long run.**
- After the introduction, comes the description of the data. The title of that section is: II. The Data. Put there what you already did in installment one.
- III. Description of the data (the part done in installment 1). This is the decomposition, which helped us get acquainted with the data.
- The next section will be: IV. ARIMA Modeling. This is what you will do for installment 2. Start this section by saying that ARIMA modeling requires that your (perhaps pretransformed-e.g., logged, etc) target variable be differenced to make it stationary. Read the instructions below for installment 2 to see what else needs to be in this section.
- The next section will be: V. Vector Autoregression. (Installment 3)
- The next section will be: V. Exponential smoothing (Installment 4)
- The next section will be: VI. Time Series regression (Installment 5)
- Finally, the last section will be: VII. Final results. Conclusions.
- Last section will be: References.
- Finally, there will be Appendices organized as follows:
  - Appendix 1. R script for section III. This contains all the code of installments I, nicely formatted. No output results. Only R code. The same R script file will be uploaded and I should be able to run it directly without any errors. I do not want to see the results or your comments. Only comments about the code are allowed, and is required. For example:
    ```
    #  We now do the additive decomposition of the data
    ```
  - Appendix 2. R script for section IV.
  - ......
  - R script for section VII.

NEW WORK FOR INSTALLMENT 2 -Section IV content:

Please, notice that everything you do for this part (2) must be done with only n-12 observations. That is, if your data set has 240 observations, do the data transformations, ACF, PACF and CCACF just with the first 228. The last 12 will be used to compare your forecasts with the actual data, to see how well your methods forecast. So, before you go on, tell R that your new data is going to be only the first 228 observations.

(a) Look at the time plots of your variables again. Define the time series by the following letters: X_t, Y_t, Z_t, F_t, where Y_t will be the label used for the target variable. In the rest of this document, Y_t refers to your target variable. Say what each of these letters denote by writing a small table with the definitions. For example: Y_t=*monthly interest rate between Jan 1010 and March 2015*. We need this notation to write the equations of your models without cluttering the pages too much. If I have to guess what X is or look beyond the page I am reading, points will be deducted. Focus from now on on your target variable. If you have pre-transformed your series, for example, you have logged, then say Y^* is the logged time series. If you don't transform, say and justify why. '

Start with that pre-transformed series, and your conclusions of Section III. For example, "in Section 3, I concluded, using decomposition, that my data has a trend and strong seasonal. I do now in this section IV an ACF to show that. The plot is in Figure...... and it reveals.......". Because I am going to use ARIMA modeling, I will regular and seasonal difference the target variable to make it stationary, qs needed." Do the necessary differencing of your target variable until you are satisfied that the ACF looks like the ACF of a stationary time series. Remember that only the ACF tells you about stationarity. You will lose points if you refer to the PACF to conclude stationarity.The lab on rooms data contains the code to do differencing. Use that as a reference. A regularly diferenced (detrended) time series Y_t on top of a logged series would be denoted as $(1-B)Y^*$. A regularly differenced and then seasonally differenced time series will be denoted as $(1-B^{12})(1-B)Y^*$. See the book for more examples. You must keep track in a table of what each final variable is, using the Backshift operator. For example, if you logged monthly Y_t, then regularly differenced, and seasonally differenced once, and that makes the data stationary, then your final data for Y_t, Section II should say at the end that you will work with:

$$(1-B^{12})(1-B)Y_t^*$$

Be aware of overdifferencing. In time series, like in any modeling, you do not want to overparameterize your models. Simple operations are always preferred to complicated ones. If a simple regular differencing does the job of leaving an ACF that cuts off at the seasonal and at the earlier lags, just use that. If you think that more than one differencing operation gives you a satisfactory ACF, please write both. You may end up identifying more than one possible model.

When you are satisfied that you have made your target series stationary (as indicated and justified by labeled and titled ACF with Figure number that follows the order of your Figures in your paper), then provide the summary of the operations using that backshift notation. If you logged, time plot of the logged data, your plot labels and titles should be clear and say what you are plotting. For example, if plotting ACF of $(1-B^{12})(1-B)X_t^*$ your plot title should have as title or label: ACF of $(1-B^{12})(1-B)Y_t^*$. Describe the final ACFs and state which r_k's are statistically significantly different from zero.

(b) Once you are satisfied with the stationarity of your target time series, denote the final pre-transformed and differenced time series by Y^{**}. For example, say

$$Y^{**} = (1-B^{12})(1-B)Y_t^*$$

And from now on you will work, for identification and modeling, with Y^{**}

(c) Do the Partial ACF for your Y_t^{**}, with the appropriate titles. For example: PACF of $(1 - B^{12})(1 - B)Y_t^*$ would be the title if you pretransformed, then regularly differenced and then seasonally differenced your target time series. If you have more than one differencing version, do the PACF for your other version as well.

(d) Using both the ACF and PACF of Y_t^{**}, identify an ARIMA(p,d,q)(P,D,Q) model for the regular part (small lag values of the time series) of your Y_t^{**}), chapter 4, 6) and the seasonal part (chapter 7). Justify your choice based on what you see in the Figures. Provide the final ACF and PACF figure (all figures must have figure numbers, so I am not repeating this requirement anymore, it should always happen).

(e) Fit your model, do the diagnostics (ACF of residuals, ljung-box test of residuals, normality of residuals, t-tests of model coefficients). This may take some trial and error, you may have to respecify your model if the residuals show autocorrelation not accounted for. The model you end up with has to be consistent with the acf and pacf you got, so do not try blind modeling. Small models with a few lags are preferred. Do not fit models with large numbers of coefficients (lags). Final plots must be provided.

(f) Once you decide which model or models are appropriate for your data, write your final conclusion in the ARIMA symbolic notation, then polynomial notation as well as the regression model notation and then forecasting model notation. For example, say that for my Y^{**} I identified an AR(1) and MA(2) part for the lower lags, and no model for the seasonal lags, because I have no seasonal autocorrelation after I differenced, y model then is in ARIMA symbolic notation

$$ARIMA(1,1,2)(0,1,0)$$

In the ARIMA polynomial notation is

$$(1 - \alpha B)(1 - B^{12})(1 - B)Y_t^* = (1 + \beta_1 B + \beta_2 B^2)w_t$$

In the regression notation, the model that you will fit is.

$$y_t^{**} = \alpha_1 y_{t-1}^{**} + \beta_1 w_{t-1} + \beta_2 w_{t-2} + w_t$$

To obtain the forecasting notation version, the computer or you will undo all the differencing. Multiply the $(1 - B^{12})(1 - B)$ to get

$y_t^* - y_{t-1}^* - y_{t-12}^* + y_{t-13}^* = \alpha_1(y_{t-1}^* - y_{t-2}^* - y_{t-13}^* + y_{t-14}^*) + \beta_1 w_{t-1} + \beta_2 w_{t-2} + w_t$

which you must leave with just y_t on the left hand side to get the forecasting model.

$y_t^* = y_{t-1}^* + y_{t-12}^* - y_{t-13}^* + \alpha_1(y_{t-1}^* - y_{t-2}^* - y_{t-13}^* + y_{t-14}^*) + \beta_1 w_{t-1} + \beta_2 w_{t-2} + w_t$

This, of course, is not the final model. Since we did a pre-differencing transformation in my example, I will have to unlog to obtain a prediction comparable to the actual raw data. So the final predicted values will be

$$y_t = e^{y_t^*}$$

This is a prediction in the same units as the original data, which is what people usually want.

(g) Once you are happy with your fitted model, forecast 12 steps ahead. Do a plot of the raw data (original units) and the forecast with the confidence bands. Start a table that has as first column the raw 12 months out-of-sample data that you took out. Then the second column will have the forecast of your ARIMA model. Put several other columns where you will put the forecasts of the other models.

(h) Compute the Root mean square error of your forecast and write it in the last row of your table under the ARIMA column.

$$\sqrt{\frac{(Y_t - \hat{Y}_t)^2}{12}}$$

(i) Upload the .pdf version of the paper again in the section provided. It should contain the corrections recommended for installment 1, fixed, all sections of installment 1, and the new section IV. The filename should be LASTNAME-ID-Installment2.pdf

(j) Upload the updated R script code for all sections. If I have to spend hours trying to figure out what each part of your code is doing, where everything is, you will lose a lot of points. Be organized, put headings such as

```
### This is the code to take the log of the interest rate series, for example..
```

(k) Upload your whole time series data in .csv format again. This file must contain all your variables. The data must be as follows: observations (time) are rows, variables are columns.