

STAT 153 Final Project

Group: Bryan Liu, Frank Qiu, Merle Behr

May 1, 2019

Summary: LaTeX is a tool used to create professional-looking documents. Below, we shall show how to insert equations, figures, and tables, as well as introduce the sections expected of the report. This template is not intended to give you a step-by-step instruction on how to do the analysis. It only helps you with LaTeX and shows you some general aspects you should take care about in your report. However, the three sections which are introduced in this template should appear exactly like this in your report, that is 1. Exploratory Data Analysis, 2. Frequency Domain Analysis, 3. ARIMA Model Selection, 4. Results. For more information about LaTeX, here are a few good references to get started

- https://www.overleaf.com/learn/latex/Learn_LaTeX_in_30_minutes
- <https://www.latex-project.org/about/>
- <https://en.wikibooks.org/wiki/LaTeX/Introduction>

In the beginning of your report, in a short summary, summarize in one or two sentences which data set you analyzed and which model you used to make predictions.

1 Exploratory Data Analysis

Here you will explore the data. Naturally, the first plot you should make is the data itself (Figure 1). Here is how you insert figures into LaTeX.

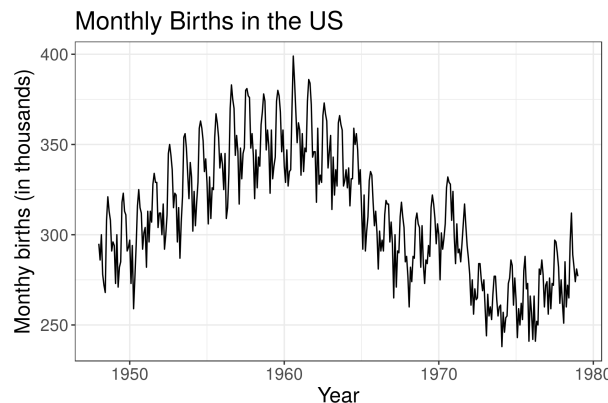


Figure 1: This is a figure. All figures should be captioned in the document. In the figure itself, axes should be properly labeled, and the font legible.

Now to apply the ARIMA models learned in this class, you must make this data stationary. More plots should follow to justify why your transforms have worked. By the end of this section, you should have a transformed time series Y_t that is stationary.

2 Frequency Domain Analysis

In this section, you will analyze the data in the Fourier domain to see if you can detect any seasonalities. Figure 2 is yet another example of how to insert plots in LaTeX. Here, we demonstrate how to use subplots in LaTeX, in case you may want to put two (or more) plots together.

Sometimes it can be a bit tricky to get figures in LaTeX placed in the position you want them to be. The command `\FloatBarrier` might be useful.

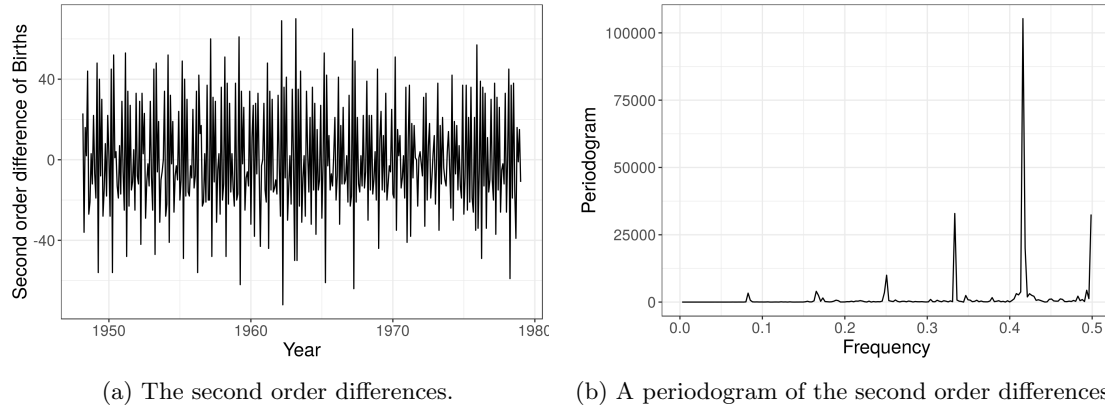


Figure 2: An example of subplots in LaTeX. Frequency analysis of the second order difference of monthly births.

3 ARIMA Model Selection

In this section, you derive an ARIMA (or SARIMA, MSARIMA, etc.) model, which you use in Section 4 to make predictions. On the way, you should compare different candidate models and do diagnostics for them. Go through the lecture materials and try your best to apply all the different approaches to find models, evaluate their accuracy, and make comparison between different models. At the end of this section you should have decide for **one single model** you want to work with for your analysis. You may also want to consider other models than ARIMA, e.g., a regression approach based on Section 2.

4 Results

Here, you will discuss the results of your model fit from Section 3. Math is very easy in Latex. Our ARIMA model is defined in equation (1). See how I referenced the equation in LaTeX?

$$(1 - \phi_1 B - \phi_2 B^2) \nabla^2 X_t = (1 - \theta_1 B) W_t \quad (1)$$

Perhaps some subsections will help the organization.

4.1 Estimation of model parameters

This might be a good time to introduce tables in LaTeX (Table 1). All tables should be captioned, accordingly.

Parameter	Estimate (s.e)
ϕ_1	-0.29 (0.06)
ϕ_2	0.068 (0.05)
θ_1	-0.99(0.007)

Table 1: These are our parameter estimates and corresponding standard errors for the ARIMA model in equation 1.

4.2 Prediction

Every group will be providing estimates of future values that are not in the dataset, and there will be one submission PER GROUP. Only the instructor will have values, and we will be evaluating your models.

Prediction File Submission:

For each of the three datasets, you'll be submitting a text file of predictions, separated by commas. For a dataset, the prediction file's name should be in the following format:

D[# of dataset]_[Group #]_[SID member 1]_[SID #2]_[SID #3]_[SID #4]_[SID #5].csv

You can view your group numbers on the posted Excel spreadsheet in bCourses. If your group doesn't have 5 members, then the remaining SID slots should be "NA". For example, if Merle, Brian, and Frank are in Group 0 submitting predictions for dataset 2, then the file name would be :

D2_0_1234_2345_3456_NA_NA.csv

**bCourses will have a sample submission for reference as well as an R script, "file_reader", for you to check if your file will be read correctly.