

# Time Series Lab

For this lab you need to upload the TS\_lab data available in the class R package.

There are **4 sets of questions** to be answered. You can get up to 100 points + bonus questions. Points are indicated next to each question.

- Format your table using stargazer
- Submit the r markdown file with all your codes.
- Bold your answers.

## 1 The policy problem

**Research question:** Does an increase in bus frequency affect bus ridership in City A?

Several cities aim to support public transportation and increase ridership. There are several ways to encourage residents to utilize public transportation more often; one of those is to modify the transit schedule so as to increase the frequency of buses in peak times and reduce waiting time.

City A has opted for this solution and starting from May 1st has implemented a new bus schedule which increases bus frequency. At the end of the year, the mayor wants to estimate whether the decision was effective and ask you to analyze one year of ridership data to test the following hypothesis:

**Hypothesis:** An increase in bus frequency has a positive effect on bus ridership in City A (i.e., the new schedule is effective).

## 2 Data

Data on ridership were provided by the local transit agency. The variable “Passengers” represents the number of daily passengers on all buses in the city (in thousands). Data were collected from January 1st to December 31st. The intervention was implemented on May 1st (day 121 is the first day of the new schedule).

Variable name	Description
Passengers	Daily passengers on the buses (in thousands)

As policy analyst, you propose to utilize a time series model to analyze the data.

## 3 Analysis

**Q1: Prepare the data for the analysis.**

- **Q1a:** Create the three variables you will need to run a time series model: Time, TimeSince, and Treatment. (5 points for each variable)
- **Q1b:** Provide a summary statistics table. (5 points)

**Q2: Run the time series model.**

- **Q2a:** Run the model and provide a result table in stargazer. (5 points + 5 points)
- **Q2b:** What does the intercept represent? (5 points)
- **Q2c:** What does the coefficient of Time represent? (5 points)
- **Q2d:** Which coefficient represents the immediate effect of the policy? (5 points)
- **Q2e:** Which coefficient represents the sustained effect of the policy? (5 points)

**Q3: Now let's look at the results more closely**

- **Q3a:** Has the new schedule increased or decreased the use of public transportation in the short term? Indicate the magnitude of the effect and whether it is statistically significant. (3 \* 3 points)
- **Q3b:** Has the new schedule increased or decreased the use of public transportation in the long term? Indicate the magnitude of the effect and whether it is statistically significant. (3 \* 3 points)
- **Q3c:** Provide a brief (1-2 statements) possible explanation for these results. (4 points)

**Q4: An important aspect of a time series model is the counterfactual.**

- **Q4a:** What is the number of passengers 100 days after the intervention? (5 points)
- **Q4b:** What is the counterfactual? Provide both its formula and estimation. (5 + 5 points)
- **Q4c:** What would the counterfactual be after 150 days? (5 points)
- **Q4d:** Are the two counterfactuals the same? Why? (3 + 5 points)

**BONUS QUESTION: Time series with a control group**

We have learned that there are threats to the validity of time series analysis. In particular, another event might have occurred at the same time of the intervention and caused the immediate and sustained effect that we observe.

A way to address this issue is to use a control group that is not subject to the intervention. This design makes sure that the effect we observe is the result of the policy intervention.

The mayor proposes to utilize city B as a control group. City B is a neighbor city with very similar characteristics to city A. Yet city B has not changed its bus schedule in the past year.

- **BQ1** Upload the data TS\_Groups\_Lab.csv from the class package. Look at the variables that are included and estimate a new time series model which includes the control group. (5 points)
- **BQ2** Interpret only the new coefficients in the model. Indicate whether they are statistically significant and what they represent. Can you confirm the results of the previous model? (5 points)