Name: Your name here
Due: 2024/11/13

Day 18 - Lab: Regression inference

Introduction

In this assignment, we will get some more practice with analyzing real data using time series regression models. To motivate our efforts, we will use average monthly CO_2 readings from the Mauna Loa observatory[1], which have been collected continuously since March of 1958. Is there evidence that the magnitude of the seasonal effect has increased over time, after accounting for the trend?

---

[1] [Link](#) to the NOAA webpage.

> **ⓘ Regression inference**
>
> In this class, we focus on two tools for making inference about regression models: the t-test and the extra sum of squares F-test. We introduce these methods briefly here, but expect to learn more about each in a regression class. The t-test, which is obtained from the `summary` output of a model, tests
>
> $$H_0 : \beta = 0 \quad \text{vs} \quad H_a : \beta \neq 0$$
>
> holding all other predictor variables constant in the model. The distribution of
>
> $$t = \frac{\hat{\beta}}{SE(\hat{\beta})}$$
>
> under the null hypothesis is t_{df}, where df denotes the degrees of freedom associated with the residual error.
>
> The extra sum of squares F-test is used to compare nested models. For example, consider two models:
>
> $$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon_i$$
> $$y_i = \beta_0 + \beta_1 x_{1i} + \epsilon_i$$
>
> We say that the second model is nested within the first model, since all parameters in the second model also exist in the first model. We can use these two models to test the following hypotheses:
>
> $$H_0 : \beta_2 = \beta_3 = 0 \quad \text{vs} \quad H_a : \text{at least one } \beta_i \text{ does not equal 0 for } i \in \{2, 3\}$$
>
> Under the null hypothesis, the distribution of the test statistic is $F_{2,df}$, which is an F distribution with two numerator and df denominator degrees of freedom, where df is the degrees of freedom associated with the residual error. To perform this test in R, we use the `anova` function: `anova(fit1, fit2)` compares models `fit1` and `fit2` in this way.

1. [1 pt] Read the Mauna Loa data in to R and filter the data to contain years prior to 1990 (not including 1990).

2. [2 pt] Plot the average CO_2 mole fraction (ppm) over time and describe the series in terms of trend and seasonality.

3. [2 pt] Fit two regression models using ordinary least squares: one that includes a time by month interaction and time squared, and one that includes time, time squared, and month (no interaction).

4. [4 pt] Assess the independence, normality, constant variance, and linearity assumptions for the interaction model.

5. [2 pt] Ignore any violations of the assumptions for now. Is the additive model nested within the interaction model? Why or why not?

6. [2 pt] Use an extra sums of squares F-test to address the research question and write up a conclusion that references: the test statistic, distribution of the test statistic under the null hypothesis, and the p-value. Ignore the violation of independence for now.

7. [4 pt] Refit both regression models using GLS and incorporate an AR(2) correlation structure in each model **using maximum likelihood as the method**. Reassess the independence assumption for the interaction model using the new fit.

8. [2 pt] Recreate the extra sums of squares F-test using the GLS models and comment on the differences.

9. [2 pt] Plot the raw time series, fitted values from the interaction model, and fitted values from the additive model on a single plot and comment on the differences.