LAB EXERCISES 4, March 31, 2023

We'll work with different data files for this lab. The files are available on the blackboard page. We'll fit and test ordinary Least Squares (OLS) regression models using the statsmodels library. Note that this is not necessarily the best library for a regression but it is easy to implement and modify for small regressions.

1. The Google Share Price data has a strong trend. We'll attempt to fit models that capture linear and non-linear effects.

   (a) Fit the following model to the whole data using OLS:

   $$y_t = \beta_0 + \beta_1 t + \epsilon_t$$

   (b) Explore the results summary and see whether the regression is significant, whether $\hat{\beta}_1$ is statistically significant. Check the R^2 value.

   (c) Compute the MSE and RMSE of the fitted model.

   (d) Fit the following model to the whole data using OLS:

   $$y_t = \beta_0 + \beta_1 t + \beta_2 \sqrt{t} + \beta_3 t^2 + \epsilon_t$$

   (e) Explore the results summary and see whether the regression is significant, whether $\hat{\beta}_i$ are is statistically significant. Check the R^2 value.

   (f) Compute the MSE and RMSE of the fitted model.

   (g) Fit the following model to the whole data using OLS:

   $$y_t = \beta_0 + \beta_1 t + \beta_2 \sqrt{t} + \beta_3 t^2 + \beta_4 \log(t) + \epsilon_t$$

   (h) Explore the results summary and see whether the regression is significant, whether $\hat{\beta}_i$ are statistically significant. Check the R^2 value.

   (i) Compute the MSE and RMSE of the fitted model.

   (j) Split the share price data into a training set (first 181 days) and a test set (days 182 to 153).

   (k) Fit the three models above to the training set and compute their MSE on the test set. What is the conclusion?

2. The Australian Beer Production data has a strong seasonality. We'll attempt to fit models that capture the effects of seasonality.

   (a) Fit the following model to the whole data using OLS:

   $$y_t = \beta_0 + \beta_1 t + \beta_2 x_{1t} + \beta_3 x_{2t} + ... + \beta_3 x_{11,t} + \epsilon_t$$

   where x_{it} are the monthly dummies for months 1 to 11 (January to November).

   (b) Explore the results summary and see whether the regression is significant, whether $\hat{\beta}_i$ are statistically significant. Check the R^2 value.

   (c) Compute the MSE and RMSE of the fitted model.

   (d) The first model has some predictors that are not statistically significant. Remove those from the model and fit a reduced model with the remaining parameters.

   (e) Explore the results summary and see whether the regression is significant, whether $\hat{\beta}_i$ are statistically significant. Check the R^2 value.

   (f) Compute the MSE and RMSE of the fitted model.

To Complete Later

1. For the Google share price data, test the following model using the train-test split from above and compute the MSE on the test set.

   $$y_t = \beta_0 + \beta_1 t + \beta_2 \log(t) + \beta_3 t^2 + \epsilon_t$$

2. For the Australian Beer Production data, implement the following train-test split: training set: first 40 months, test set: last 16 months. Compare the two models by fitting them on the train data and computing their MSEs on the test set.