

Assignment 2

You will do this assignment in groups of up to 4. Once you have formed a group you can email me (hansh@ru.is) and get your group's data.

You get two data files with simulated data representing monthly sales at a retail company over a period of 10 years and some potential predictors. The variables in the files are as follows:

Wide data:

purchase_power: An index of average purchase power in the country where the retail company operates. Set to 100 on time 1. Changes are percentage changes from time 1, i.e. if the index is at 120 that means purchase power is 20% greater than at time 1. If the index is at 95 purchase power is 5% lower than at time 1.

purchase_power_LAG1: The purchase power index lagged by one month.

purchase_power_LAG2: The purchase power index lagged by two months.

purchase_power_LAG3: The purchase power index lagged by three months.

consumer_sentiment: An index of a measure of consumer sentiment in the country where the retail company operates. Set to 100 at time 1.

consumer_sentiment_LAG1: The consumer sentiment index lagged by one month.

consumer_sentiment_LAG2: The consumer sentiment index lagged by two months.

consumer_sentiment_LAG3: The consumer sentiment index lagged by three months.

sales: An index of total sales. Set to 100 at time 1.

sales_LAG1: The sales index lagged by one month.

sales_LAG2: The sales index lagged by two months.

sales_LAG3: The sales index lagged by three months.

month: Calendar month.

month_num: Calendar month represented as number. January is 1, February is 2, March is 3 etc.

time: Months in data set numbered sequentially.

year: Years in data set numbered sequentially. Note that the time series begins in October of year 1 and ends in September of year 11.

Long data:

time: Months in data set numbered sequentially.

month_num: Calendar month represented as number. January is 1, February is 2, March is 3 etc.

year: Years in data set numbered sequentially.

name: Index represented in each row. Takes the following values: *purchase_power*, *consumer_sentiment* and *sales*.

value: The value of the index represented in each row. Note that the long data does not have lagged values.

Tasks:

1) Make a line chart showing all three indexes over time on the same plot. You can do this in BlueSky under Graphics-> Line Charts-> Points

Looking at the plot, do you see any indication that one variable is leading any other?

2) Fit a simple regression model where time predicts sales. Make sure to name the model (in the "Enter Model Name" field). Save the residuals to data. This can be done under Model Evaluation -> Add Statistics To Dataset. You then need to check the .resid box.

Make a line chart of the residuals (with time on the x-axis).

Install the lmtest package (instructions on slides). Use the dwtest function to check for first order autocorrelation in the residuals. This can be done by entering `dwtest(regressionModelName)` into the R editor.

Use the Dickey-Fuller test to test for stationarity in the residuals. (Instructions on slides)

3) Fit the appropriate regression models and compare them to determine if a) *purchase_power* and b) *consumer_sentiment* Granger-cause sales (use the three lagged values provided in the data).

Also test if one Granger-causes the other.

Return the assignment in the form of a short report. Show the appropriate plots and tables from the BlueSky output and explain as necessary.

