**Assignment 4: Time Series Analysis (100 points)**

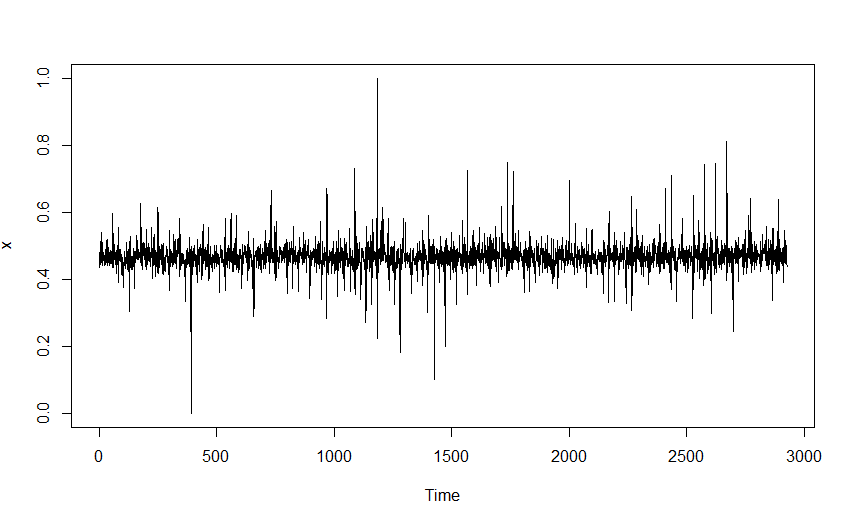
**Student Name:**

**Purpose:** To perform time series analysis to predict future customer sentiment on Twitter

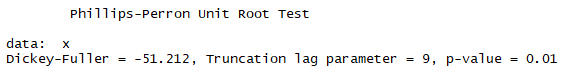
**Description:** The objective of this study is to find the right time series model to make predictions about the future customer sentiment in Twitter network of AT&T. The data contains the average hourly sentiment of tweets customers sent to AT&T’s Twitter handle.

**Instructions:** Please follow these steps:

1. In Canvas, navigate to Assignments and then Assignment4
2. Download and save the data set ATT\_Twitter.csv
3. Read the file: data <- fread("ATT\_Twitter.csv", sep=",", header=T, strip.white = T, na.strings = c("NA","NaN","","?"))
4. Use packages “forecast”, “timeSeries”, and “rugarch” to answer the following questions:
   1. **(5 points)** Paste the plot of the time series in the space below:

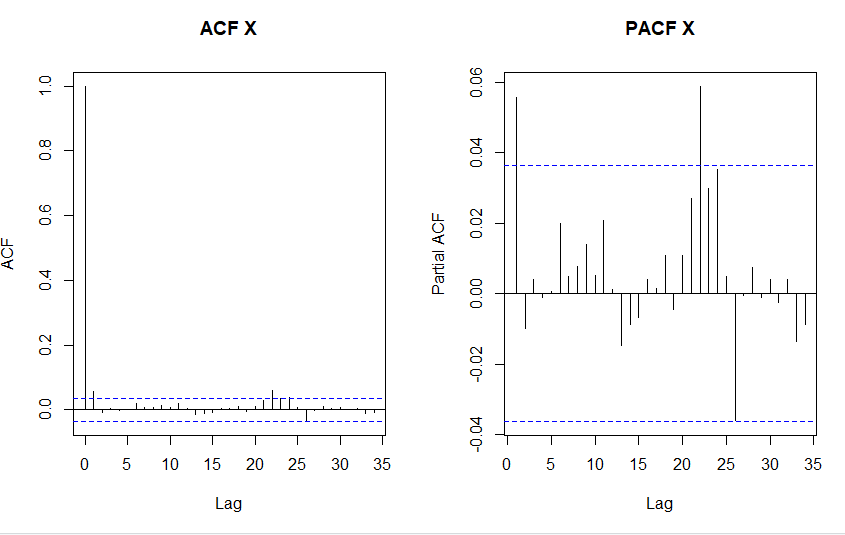


* 1. **(10 points)** Determine if this time series is a random walk process?



The p-value is 0.01. This suggests that we can reject the null hypothesis that the variable contains a unit root. Therefore, this series is not a random walk.

* 1. **(20 points)** Use ACF and PACF to determine if the times series has any MA or AR process. If so, what is (are) the order(s)? Please paste the ACF and PACF plots in the space below:

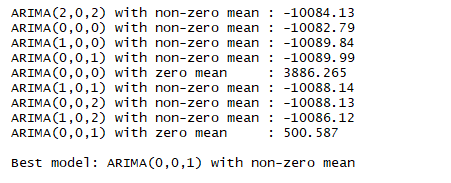


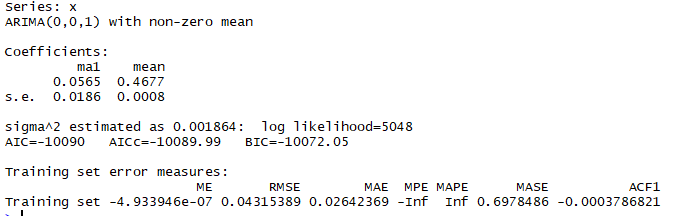
It’s close but according to the ACF plot above, my interpretation is that none of the lags are significantly correlated. Therefore, I can assert that the data could be white noise. In other words, this data cannot be distinguished from white noise.

Examining the PACF, it appears that there are some significant partial correlations with the time series after approximately 22 hours. This correlates with the ACF as well.

I think at this point, we can say the time series has no MA or AR processes but we need to note the discrepancies in the plots above. Further monitoring and analysis are necessary to reach a stronger conclusion and to fully investigate patterns in the data. We might see a partial correlation every 22 hours.

* 1. **(5 points)** Use auto.arima function to determine the best model ARIMA model for this data. What is the best order?





The best model is ARIMA(0,0,1) with a non-zero mean. The best order is 0,0,1.

1. Use “rugarch” package to build ARCH and GARCH models. Use “sGARCH” in the specification.
   1. **(5 points)** First use armaOrder(0,0). What are the values of AIC and BIC?

AIC = -3.4473

BIC = -3.4392

* 1. **(5 points)** Now use armaOrder(0,1). What are the values of AIC and BIC?

AIC = -3.4451

BIC = -3.4348

1. Use “rugarch” package to build ARCH and GARCH models. Use “apARCH” in the specification.
   1. **(5 points)** First use armaOrder(0,0). What are the values of AIC and BIC?

AIC = -3.6601

BIC = -3.6478

* 1. **(5 points)** Now use armaOrder(0,1). What are the values of AIC and BIC?

AIC = -3.6672

BIC = -3.6529

* 1. **(10 points)** Using scholarly articles, explain what is the main difference between apARCH (asymmetric power ARCH) and simple GARCH models?

Generalized autoregressive conditional heteroskedasticity (GARCH) treats heteroskedasticity as a function of time and doesn’t take into account random periods of risk or volatility.

<https://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.15.4.157>

However, risky times are not spread evenly across time and cannot be forecasted as a simple function of time. Asymmetric power ARCH (APARCH) leverages negativity and closer periods of time, t. This is because volatility tends to bunch and happen at once as opposed to evenly over time.

<https://klevas.mif.vu.lt/~danas/TS/WurtzEtAlGarch.pdf>

However, the original authors of a paper on APARCH have released an edit questioning whether or not APARCH works better than simple GARCH modeling.

<https://www.aeaweb.org/articles?id=10.1257/jep.15.4.157>

1. **(30 points)** Based on your analysis in 5 and 6, which model specification you would choose for predicting the future values? How the values of AIC and BIC influence your decision?

Time series have heteroscedastic variances. Although some variances increase gradually over time, others are seemingly more sporadic or have shorter periods of volatility. ARCH and GARCH models seek to “explain” these shorter and more variable variances.

These models are fitted to a given time series. However, it is important to understand the goodness of fit of these models. This can be done with the following two methods, Akaike’s Information Criterion (AIC) and Bayesian Information Criterion (BIC). The main difference between the two is that BIC penalizes the number of parameters more strongly than AIC does. Typically, the lower number AIC and BIC values are indicative of the “best fit” model.

Given what we know about AIC and BIC, our best model for predicting future values is armaOrder(0,1) using an apARCH in the specification. The values are as follows:

AIC = -3.6672

BIC = -3.6529