Econ 4140 Assignment 1

The files TESLA.xlsx txt (excel format) and TESLA.txt (tex format) contain daily data on prices of TESLA stock from 2019/01/02 to 2021/12/30. Use the SAS code provided for Assignment 1 or your own code to perform the following analysis.

Examine the daily close prices and answer the following questions:

1. Define the random walk process.

1.1 General Definition:

Let $Z_1, Z_2, \ldots$ be independent and identically distributed, with mean μ and standard deviation σ. Let S_0 be an arbitrary starting point, which implies.

$$S_t = S_0 + Z_1 + \cdots + Z_t, \quad \text{where } \ t \geq 1$$

S_t is the value of the random walk and Z_t are it's steps. If these steps are $\sim N(\mu, \sigma^2)$ then it is called a **normal random walk**

How do you interpret and visualize the variables of the random walk?

> As mentioned before S_0 is the starting point of the random walk.
> μ is known as the **drift** of the random walk and determines the overall direction.
> σ is called the **volatility** and determines how much the random walk fluctuates around it's conditional mean $(S_0 + \mu t)$.

1.2 Geometric Random Walks:

A geometric random walk uses the log prices of a stock to gauge returns rather than regular prices. While intuitively it seems bizarre, it ends up being a more accurate measure of returns and allows for more accurate forecasting.

> There are 2 main reasons on why geometric random walks are preferred to regular random walks.
> 1. Geometric random walks imply nonnegative prices, where regular random walks do. Prices in real life aren't negative.
> 2. Geometric random walks give net returns that are ≥ -1. In real life you cannot possibly lose more than 100% of your investment.

Mathematically, the lognormal geometric random walk hypothesis states that the single-period log returns: $r_t = log(1 + R_t)$, are independent and identically distributed ((i.i.d) or $\sim N(\mu, \sigma^2)$.

This is because:

$$\begin{aligned} 1 + R_t(k) &= (1 + R_t) \ldots (1 + R_{t-k+1}) \\ &= \exp(r_t) \ldots \exp(r_{t-k+1}) \\ &= \exp(r_t + \cdots + r_{t-k+1}) \end{aligned}$$

therefore:

$$\begin{aligned} \log\{1 + R_t(k)\} &= r_t + \cdots + r_{t-k+1} \\ \frac{P_t}{P_{t-k}} &= 1 + R_t(k) = exp(r_t + \cdots + r_{t-k+1}) \end{aligned}$$

Because we are given Tesla's daily stock prices in the data we can use this relation to find the returns.

1.3 Interpretation in terms of TESLA's prices and the code:

As we can see from line 4 in the code, TESLA's returns are defined as the difference in log prices. This is congruent with our statement above in the section about geometric random walks ($\frac{P_t}{P_{t-k}} = 1 + R_t(k) = exp(r_t + \cdots + r_{t-k+1})$).

> Why is this?

Well we know from our logarithm rules that:

$$\log(\frac{m}{n}) = \log(m) - \log(n)$$

Because $\log(1 + R_t(k)) = r_t$ finding our returns requires us to take the log of both sides.

$$\log(\frac{P_t}{P_{t-k}}) = \log(1 + R_t(k))$$

$$r_t = \log P_t - \log P_{t-k}$$

Which gives rise to line 4 in the SAS code:

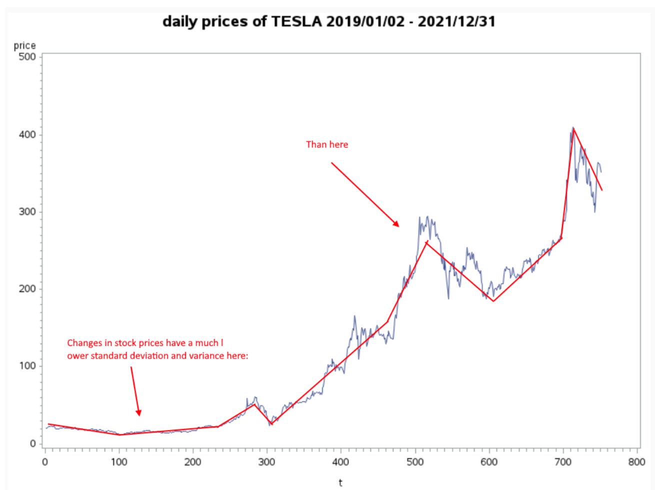
```
data tesla;
infile '/home/u64036772/TESLA.txt';
input price;
return = dif(log(price));
t=_n_;
```

So while we cannot confirm for sure as of yet that Tesla's returns do in fact follow a random walk instead of being a stationary process, we can gain the intuition that at the very least it would follow a geometric random walk rather than a regular random walk.

2. Does the time series of daily TESLA prices display a trend?

When we ask "does this random variable display a trend?", what you are actually asking is if the random variable is stationary or not.

A stationarity process is one whose probability distribution does not change when its start is shifted in time. Consequently it's parameters such as it's mean, standard deviation and variance change if the process is NOT stationary. When we look at the graph of TESLA prices, we can see that the mean and standard deviation appear to change over time.



When we go to SAS we can input an additional section of code:

An AUTOREG model on **prices** with lag 1:

```
proc autoreg;

model price =/nlag = 1;

run;
```

With this we can use ϕ to check to see if our process is stationary or not:
If $\phi < 1$ then $y_1, \ldots$ is a weakly stationary process.

If $\phi = 0$ then it is a white noise process (stationary) -

If $\phi < 1$ then it is an AR(1) process (stationary)
If $\phi \geq 1$ then it has a trend, and acts like a random walk.

Our proc autoreg on prices gives an estimate of the ϕ which is:

Estimates of Autoregressive Parameters			
Lag	Coefficient	Standard Error	t Value
1	-0.994934	0.003673	-270.84

Except SAS interprets this number in AUTOREG as negative, but the way we have defined it is actually positive.

So phi is actually $\phi = 0.994934$
Which is very close to 1, but not actually 1.

So while the graph makes it look like the process isn't stationary and actually a random walk, mathematically it has to be because our estimate of $\phi < 1$.

3. Given your answer to the first question, do TESLA prices behave like a stationary process or rather like a random walk?

As we found in question 2, TESLA prices appear to behave like a random walk, but in reality it is a AR(1) stationary process because $\phi < 1$.

4. Does the return series, defined as $return(t) = log(p_{(t)}) – log(p_{(t-1)})$ have a trend?

As stated earlier if a time series fits all the assumptions of weak stationarity then it is said to have no trend. The easiest way again is to look at ϕ.

Our output from SAS:

Estimates of Autoregressive Parameters			
Lag	Coefficient	Standard Error	t Value
1	-0.007354	0.036563	-0.20

Therefore $\phi = 0.007354$,

As a result we can conclude that our return series is also an AR(1) process and is stationary.

5. What are the differences between the behavior of prices and returns?

The behavior of returns are directly dependent on the behavior of prices, because our return is defined as $r_t = \log(P_t) - \log(P_{t-1})$, so if we were to change the behavior of prices, it would change the behavior of returns.

But this doesn't mean that there aren't any differences between the two, in fact there are many.

For example, the conditional probability distribution of returns is completely different than the conditional probability distribution of prices. This means that they both have different location, scale and shape parameters including mean, variance, kurtosis, and quantiles.

Despite this, they are both still AR(1) processes, and as a result mathematically exhibit the same properties of:

1. $E(y_t) = \mu \quad \forall t$
2. $\gamma(0) = Var(y_t) = \frac{\sigma_\epsilon^2}{1-\phi^2} \quad \forall t$
3. $\gamma(h) = Cov(y_t, y_{t+h}) = \frac{\sigma_\epsilon^2 \phi^{|h|}}{1-\phi^2} \quad \forall t$
4. $\rho(h) = Corr(y_t, y_{t+h}) = \phi^{|h|} \quad \forall t$

So while their actual means wont be the same ($\mu_{r_t} \neq \mu_{p_t}$), for both time series processes these 4 equations will hold true.

A great example of this is through the examination of volatility clustering. While the actual variance/standard deviations of each series are different, increases in variance/standard deviations in prices directly increases variances/standard deviation in returns.

6. Define “volatility clustering”.

Volatility clustering is where time series financial market data exhibits changing periods of high and low volatility (rather than a constant volatility).

Volatility clustering can be detected by looking for auto-correlation in the mean-centered squared residuals.

Page 406 of "Statistics and Financial Engineering" by David Ruppert provides a great figure showing economic data with time-varying volatility (volatility clustering). **(Fig 14.1)**

406 14 GARCH Models

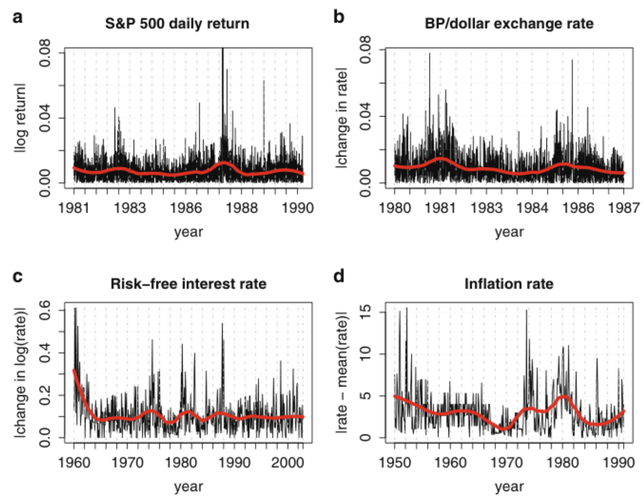
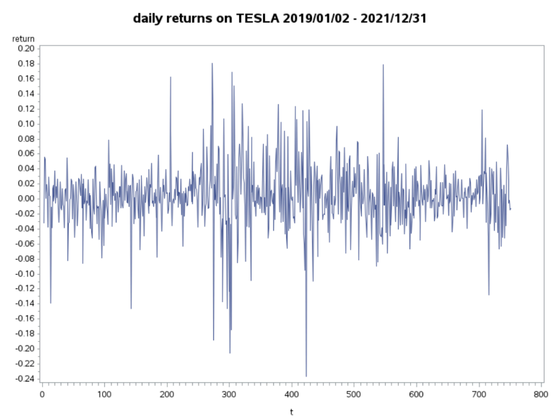


Fig. 14.1. *Examples of financial markets and economic data with time-varying volatility: (a) absolute values of S&P 500 log returns; (b) absolute values of changes in the BP/dollar exchange rate; (c) absolute values of changes in the log of the risk-free interest rate; (d) absolute deviations of the inflation rate from its mean. Loess (see Section 21.2) smooths have been added in red.*

Furthermore, we can also see volatility clustering in the actual graph of our time series of returns.



Our returns from 0-250 experience a much lower level of volatility than 250-500, and a muich higher level than 500-750.

7. Comment on the patterns in return and prices between 2020/01/29 and 2020/03/13 - what do you observe?

The volatility of TESLA prices and returns increased significantly, the distribution exhibits volatility clustering.

8. Are the TESLA returns normally distributed? Explain and describe all evidence from the output provided by the summary statistics and figures (histogram, qqplot, quantiles)

No, TESLA's returns are not normally distributed for multiple reasons.

1. **All 4 of our tests for normality have a p value of** < 0.05

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.929386	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.086314	Pr > D	<0.0100
Cramer-von Mises	W-Sq	2.019849	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	11.81865	Pr > A-Sq	<0.0050

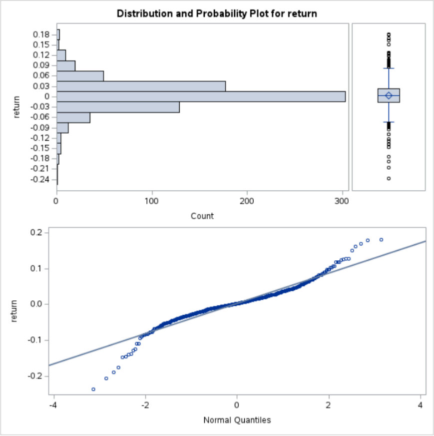
This allows us to reasonably assume that our assumption of normality doesn't hold true. Why? When we do our assumptions of normality we test it against the null hypothesis that returns are normally distributed versus the alternate hypothesis that they aren't.

$$H_0 : r_t \sim N(\mu, \sigma^2)$$

$$H_A : r_t \neq N(\mu, \sigma^2)$$

When we go to do our tests for normality, we find that all 4 of our p values for our tests are < 0.05 which means that we can reject the null hypothesis(H_0) that $r_t \sim N(\mu, \sigma^2)$, and do not reject the alternative hypothesis$(H_A : r_t \neq N(\mu, \sigma^2))$.

2. **TESLA's distribution and probability plot for returns displays multiple inconsistencies with a normal distribution and probability plot**



As you can see in our distribution plot, tesla's lower tail is much longer than what a normal distribution would have, and that the distribution is skewed and asymmetric.

Furthermore, our probability plot testing our quantiles shows that we have much fatter tails than a normal distribution. What metric could quantify differences in tails from normality? - Excess Kurtosis

3. **The distribution has excess kurtosis and significant skew (statistically)**

Moments			
N	750	Sum Weights	750
Mean	0.00378061	Sum Observations	2.83546043
Std Deviation	0.04233713	Variance	0.00179243
Skewness	-0.2953759	Kurtosis	4.83367284
Uncorrected SS	1.3532515	Corrected SS	1.34253172
Coeff Variation	1119.8479	Std Error Mean	0.00154593

SAS reports "Kurtosis" as "Excess Kurtosis" so the actual kurtosis is 7.833 whereas our excess kurtosis is what is shown in the plot (4.833).

Additionally, the distribution is skewed and as we know, normal distributions are completely symmetric.

9. Write the formula of the AR(1) process.

Which AR(1) process? The general one, for prices or for returns?

General Formula of an AR(1) process

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + \epsilon_t$$

$$\text{where } \epsilon_t \sim WN(0, \sigma_\epsilon^2)$$

and

$$Y_t - \hat{Y}_t = \epsilon_t$$

But this can be rewritten more simply where Y_t is a function of Y_{t-1}

$$Y_t = (1-\phi)\mu + \phi Y_{t-1} + \epsilon_t$$

AR(1) Prices:

$$Y_t = (1-\phi)\mu + \phi Y_{t-1} + \epsilon_t$$

$$B_0 = (1-\phi)\mu$$

Parameter Estimates					
Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	145.8897	36.9318	3.95	<.0001

Estimates of Autoregressive Parameters			
Lag	Coefficient	Standard Error	t Value
1	-0.994934	0.003673	-270.84

Pulling our coefficients directly from the data we get:

$$Y_t = 145.8897 + 0.994934 Y_{t-1} + \epsilon_t$$

AR(1) Returns:

$$Y_t = (1-\phi)\mu + \phi Y_{t-1} + \epsilon_t$$

As we know $B_0 = (1-\phi)\mu$

But we dont even need to calculate B_0 as it's given to us.

Parameter Estimates					
Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	0.003781	0.001546	2.45	0.0147

Estimates of Autoregressive Parameters			
Lag	Coefficient	Standard Error	t Value
1	-0.007354	0.036563	-0.20

$$Y_t = 0.003781 + 0.007354 Y_{t-1} + \epsilon_t$$

10. Explain the computer output and describe all evidence along the following lines:

10.1 what are the estimated marginal mean and variance of returns? Is the mean return statistically significant? – write the statistic for testing the hypothesis: “mean return = 0”

The estimated marginal mean and variance are:

$\mu = 0.003781$

$\sigma^2 = 0.00179$

As shown in the table below:

Basic Statistical Measures			
Location		Variability	
Mean	0.003781	Std Deviation	0.04234
Median	0.002698	Variance	0.00179
Mode	.	Range	0.41796
		Interquartile Range	0.03981

As for the mean return being statistically significant, we can use Student's t to test it. If our $Pr > |t|$ is less than 0.05, then we can reject the null hypothesis.

Mathematically we write our mean tests as:

$$H_0 : \mu = 0$$
$$H_A : \mu \neq 0$$

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	2.445522	Pr > |t|	0.0147
Sign	M	35	Pr >= |M|	0.0117
Signed Rank	S	18242.5	Pr >= |S|	0.0021

and from the table we can see that it is in fact less than 0.05, therefore we reject the null hypothesis that our mean is zero and do not reject the alternative hypothesis.

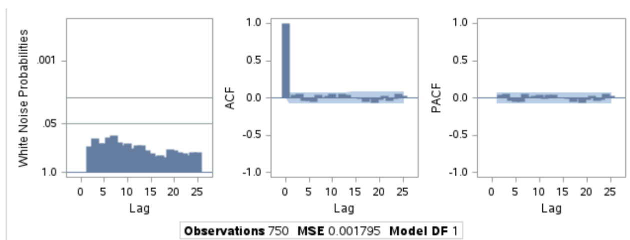
So, we can conclude that our results are in fact statistically significant and that the mean does not equal 0.

10.2 Are the returns serially correlated? Is the autoregressive coefficient of the estimated AR(1) model statistically significant? What is the estimated variance of the error term of the AR(1) model?

Yes returns are serially correlated, but by a very small fraction. This is because our ϕ on returns is quite small ($\phi = 0.007354$), and so our returns have little memory.

You can look for the autocorrelation in residuals to find evidence against the statistical significance of an AR(1) model.

To test for residual autocorrelations use the test bounds, anything outside of the test bounds (with the exception of $t_0 = 1$) is evidence against the AR(1) assumption.



So when we go to look at our white noise probabilities (which is the SAME thing as saying $H_0 : \phi = 0$, our null hypothesis that our AR(1) process on returns is $\sim WN(0, 1)$) we find that NONE of our residual autocorrelations are outside of the test bounds.

As a result we can conclude that our AR(1) model is in fact statistically significant.

10.3 What is the efficient market hypothesis?

The efficient market hypothesis states that share prices reflect all available information in the economy at the given time. There are different forms of the efficient market hypothesis namely strong, semi-strong and week.

The strong form efficient market hypothesis says that all information including information not publicly known (insider information) is completely accounted for in a stock price (past, current and future information)

The semi-strong form efficient market hypothesis states that only current and past information is accounted for in a stock price

The weak form efficient market hypothesis states that only past information is included in a stock's price.

How does this tie into our study of TESLA's returns?

The efficient market hypothesis predicts that log returns on stocks will be white noise.

10.4 Is your AR(1) estimation result consistent with the efficient market hypothesis? Explain why yes or not.

Yes our AR(1) estimation result has to be consistent with at least the **weak-form efficient market hypothesis** because our model is based off of the companies past stock return data.