Mattias Berg, Colton Cunov

Financial Timeseries: Project 1

Theoretical part

1.)

Let $\mu > 0$ and $\sigma^2 > 0$ and let $Z \sim \mathrm{WN}(\mu, \sigma^2)$. Let then Y be the process defined by $Y_t = \sum_{j=0}^{q} \theta_j Z_{t-j}$ for some coefficients $\theta_i, ..., \theta_q \in \mathbb{R},\ q \in \mathbb{N}$ with $\theta_0 = 1$. As we will see in the lectures, is called a moving average process of order .

a.) Show that for any $(t, h) \in \mathbb{Z}, E(Y_t) = E(Y_{t+h})$

Here we are showing that the process is stationary.

$r = t,\ s = t + h$

$E(Z_r) = E(Z_s) = \mu$

$$E\left(Y_r\right) = E\left(\sum_{j=0}^{q} \theta_j Z_{r-j}\right) = E(\theta_0 Z_r + \theta_1 Z_{r-1} + ... + \theta_q Z_{r-q}) = E(\theta_0 Z_r) + E(\theta_1 Z_{r-1}) + ... + E(\theta_q Z_{r-q}) = \mu\,(\theta_0 + \theta_1 + ... + \theta_q)$$
$$E\left(Y_s\right) = E\left(\sum_{j=0}^{q} \theta_j Z_{s-j}\right) = \ ... \ = \mu \sum_{j=0}^{q} \theta_j = E(Y_r)$$

b.) Show that $(t, s, h) \in \mathbb{Z}^3$, $Cov(Y_t, Y_{t+h}) = Cov(Y_s, Y_{s+h})$;

Here we are showing that the covariance is only dependent on the lag *h*.

$\mathrm{Cov}(Y_t, Y_{t+h}) = E[(Y_t - \mu_Y(t))(Y_{t+h} - \mu_Y(t+h))] = E(Y_t Y_{t+h}) - \mu_Y(t)\mu_Y(t+h)$

$\mathrm{Cov}(Y_s, Y_{s+h}) = E[(Y_s - \mu_Y(s))(Y_{s+h} - \mu_Y(s+h))] = E(Y_s Y_{s+h}) - \mu_Y(s)\mu_Y(s+h)$

Since; $E(Y_t) = E(Y_{t+h}) = \mu_Y(t) = \mu_Y(t+h) := \mu_Y$

$\mathrm{Cov}\left(Y_t, Y_{t+h}\right) = E\left(Y_t Y_{t+h}\right) - {\mu_Y}^2$

$\mathrm{Cov}\left(Y_s, Y_{s+h}\right) = E\left(Y_s Y_{s+h}\right) - {\mu_Y}^2$

where

$$E(Y_t Y_{t+h}) = E\left[\left(\sum_{j=0}^{q} \theta_j Z_{t-j}\right)\left(\sum_{k=0}^{q} \theta_k Z_{t+h-k}\right)\right] = \sum_{j=0}^{q}\sum_{k=0}^{q} \theta_j \theta_k E(Z_{t-j} Z_{t+h-k}) = \mu^2 \sum_{j=0}^{q}\sum_{k=0}^{q} \theta_j \theta_k = E(Y_s Y_{s+h})$$

so $\mathrm{Cov}\left(Y_t, Y_{t+h}\right) = E\left(Y_t Y_{t+h}\right) - {\mu_Y}^2 = \mathrm{Cov}\left(Y_s, Y_{s+h}\right) = E\left(Y_s Y_{s+h}\right) - {\mu_Y}^2$

c.) Show that Y is stationary and give its autocovariance function.

Weak stationarity: $\mu_Y(t) = c \ \forall \ t \in \mathbb{Z}$ and $\gamma_Y(r,s) = \gamma_Y(r+h, s+h) \ \forall \ (r,s,h) \in \mathbb{Z}^3$

As shown in **a)**, $E(Y_t) = E(Y_{t+h}) = \mu_Y(t) = \mu_Y(t+h)$ is constant.

As shown in **b)**, the covariance function only depends on the lag *h*. With $(r,s) \leftrightarrow (r+h, s+h)$ we see a shift of *h* and lag of $s-r$. In **b)** we had $(t, t+h) \leftrightarrow (s, s+h)$ which has a shift of $s-t$ and a lag of *h*.

ACF:$\gamma_Y(h) := Cov(Y_{t+h}, Y_t)$

From **b)**

$$\mathrm{Cov}\left(Y_t, Y_{t+h}\right) = \mu^2\left(\sum_{j=0}^{q}\sum_{k=0}^{q}\theta_j\theta_k\right) - \mu_Y^2 = \mu^2\left(\sum_{j=0}^{q}\sum_{k=0}^{q}\theta_j\theta_k\right) - \left(\mu\sum_{j=0}^{q}\theta_j\right)^2 =$$
$$= \mu^2\left(\sum_{j=0}^{q}\sum_{k=0}^{q}\theta_j\theta_k - \left(\sum_{j=0}^{q}\theta_j\right)^2\right) = \mu^2\left(\sum_{j=0}^{q}\sum_{k=0,k\neq j}^{q}\theta_j\theta_k\right)$$

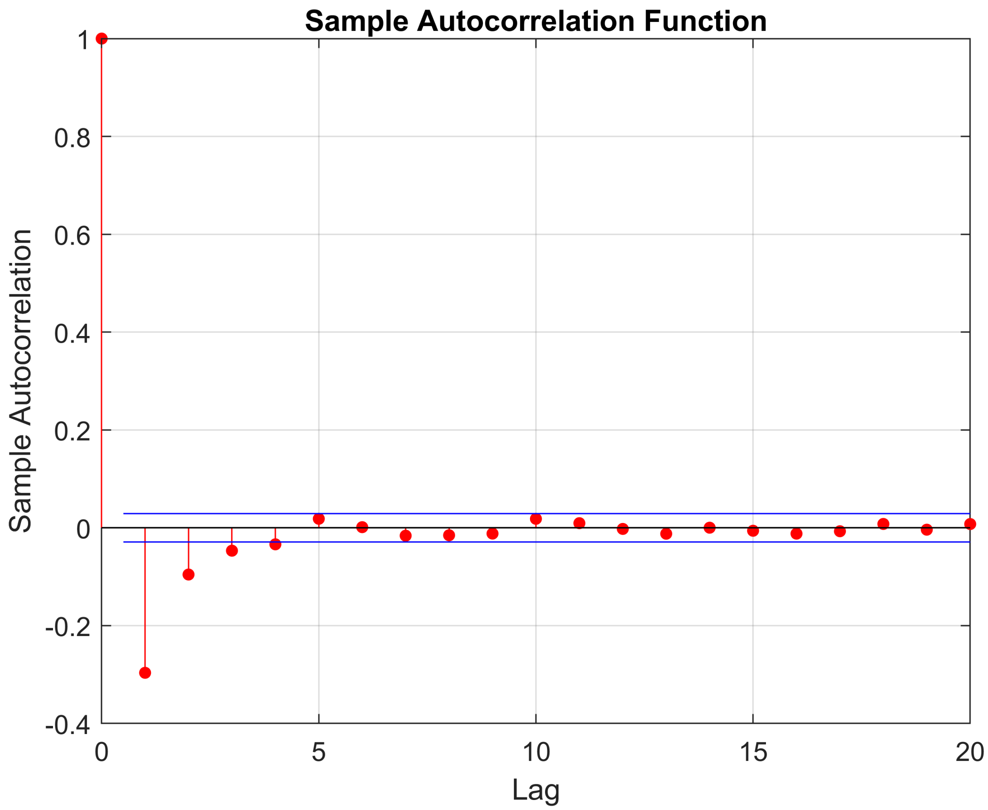
d.) Prove that if Z is a Gaussian process, then Y_t is independent of Y_{t+h} for any $t \in Z$ and $|h| > q$.

If Z is Gaussian then it is necessarily iid, and consequently Y_t is for $|h| > q$. When $|h| \leq q$, we have $|q - h + 1|$ Z's that appear in both Y_t and Y_{t+h}, implying the realization of Y_t influences the probability distribution of Y_{t+h} and therefore non-independence. This is not the case for $|h| > q$, when all Z's that appear in Y_t and Y_{t+h} are unique realizations of $Z \sim \mathrm{WN}(\mu, \sigma^2)$.

Practical part

2.)

```
clf
d = readtable('intel.csv');
X_miss = d.VolumeMissing;
Y_miss = log(X_miss(2:end)) - log(X_miss(1:end-1));
autocorr(Y_miss, 20)
```



We look at the ACF to determine q via the bounds +/- 1.96/srqt(n) (95% confidence interval) gain from the autocorr function plot above. Choosing q to be the last value of h that is outside the confidence interval. Assuming that all h after is IID noise, as they are all within the confidence interval (h=5...20). From this we assume that a reasonable value of q is 4.

3.)

```
q = 4; % q value gained in question above
acf = autocorr(Y_miss, 20);
M = find(isnan(Y_miss)); % array with index of all the missing data
```

We now want to solve **a** in;

$$\begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(q-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(q-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(q-1) & \gamma(q-2) & \cdots & \gamma(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_q \end{bmatrix} = \begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(q) \end{bmatrix}$$

with γ being the autocorr values we got earlier.

We set **n** to be **q**, as values above are IDD noise, and create the autocovariance matrix;

```
acfMatrix = toeplitz(acf(1:q));
```

Solve the system to get $a_1 \cdots a_q$

```
a = linsolve(acfMatrix,acf(2:q+1));
```

We now solve for the missing values by using $b_t^l(Y^q) = a_0 + a_1 Y_{t-1} + a_1 Y_{t+1} + \cdots + a_q Y_{t-q} + a_q Y_{t+q}$. Note that a_0 will always be 0, as $\mu = 0$.

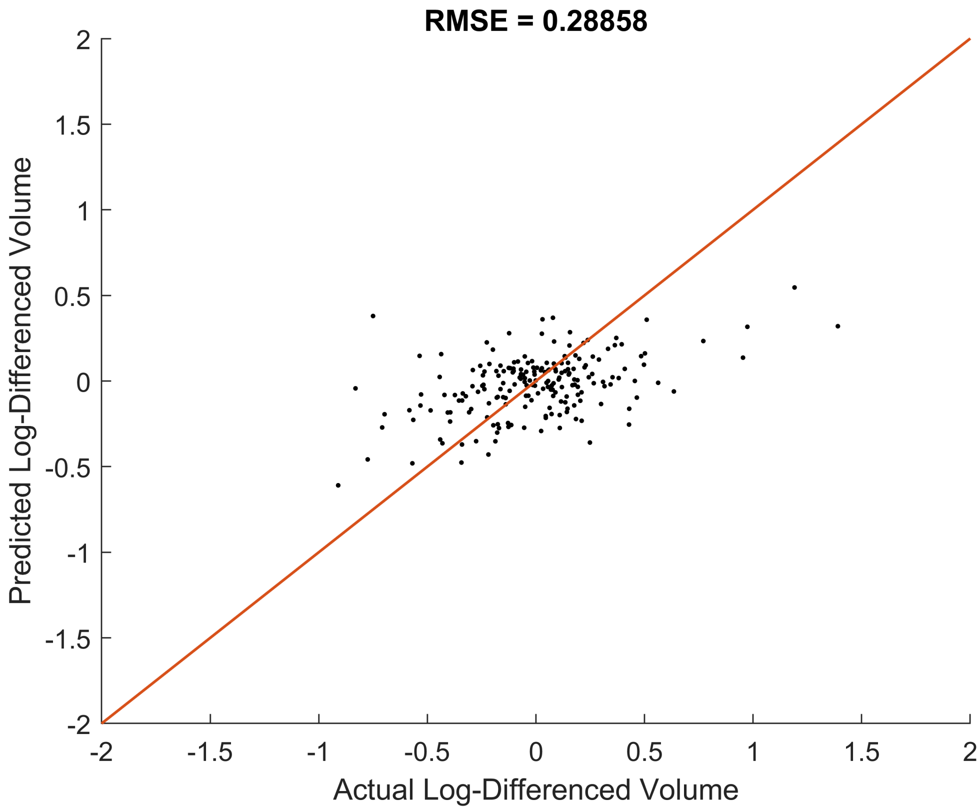
```
Y_solved=Y_miss;
for index = 1:length(M)
    y_temp = 0;
    %Looping over all indexes (s) in max(1,t-q) ≤ s ≤ min(N,t+q)
    minTime = max([1, M(index)-q]);
    maxTime = min([length(Y_solved), M(index)+q]);
    for s = minTime:maxTime
        %Check that data is not a member of M.
        if ~ismember(s, M)
            y_temp = y_temp + a(abs(s-M(index)))*Y_solved(s);
        end
    end
    %Set the calculated value inplace of the missing value (NaN)
    Y_solved(M(index))=y_temp;
end
```

4.)

The calculated RMSE for $\hat{Y}$ can be seen in the plot generated by the code below.

```
X = d.Volume;
Y = log(X(2:end)) - log(X(1:end-1));

RMSE = sqrt(mean((Y(M) - Y_solved(M)).^2));
scatter(Y(M),Y_solved(M),'k.');
hold on
plot(-3:3,-3:3,'LineWidth',1)
hold off
xlim([-2 2])
ylim([-2 2])
title(['RMSE = ' num2str(RMSE)])
xlabel('Actual Log-Differenced Volume')
ylabel('Predicted Log-Differenced Volume')
```



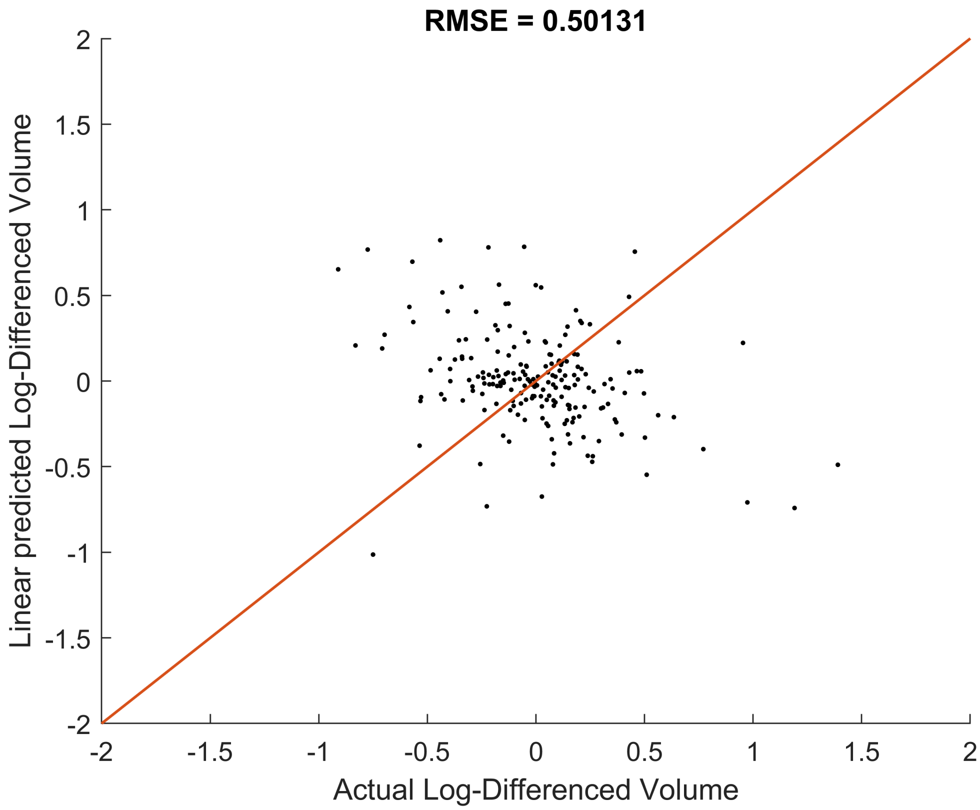

```
MAPE = mean(abs((Y(M) - Y_solved(M))./Y(M)));
disp(['Mean absolute % error = ' num2str(MAPE)])
```

```
Mean absolute % error = 1.5782
```

The calculated RMSE for $\check{Y}$ can be seen in the plot generated by the code below.

```
Y_linear = fillmissing(Y_miss,'linear');

RMSE_linear = sqrt(mean((Y(M) - Y_linear(M)).^2));
scatter(Y(M),Y_linear(M),'k.');
hold on
plot(-3:3,-3:3,'LineWidth',1)
hold off
xlim([-2 2])
ylim([-2 2])
title(['RMSE = ' num2str(RMSE_linear)])
xlabel('Actual Log-Differenced Volume')
ylabel('Linear predicted Log-Differenced Volume')
```



As seen in the two plots above, our own prediction ($\hat{Y}$) gives a better result than using a simple linear prediction ($\check{Y}$) of the data.