Time Series Project 1

Stefan Eng and Franz Hartleitner

April 2019

Problem 1
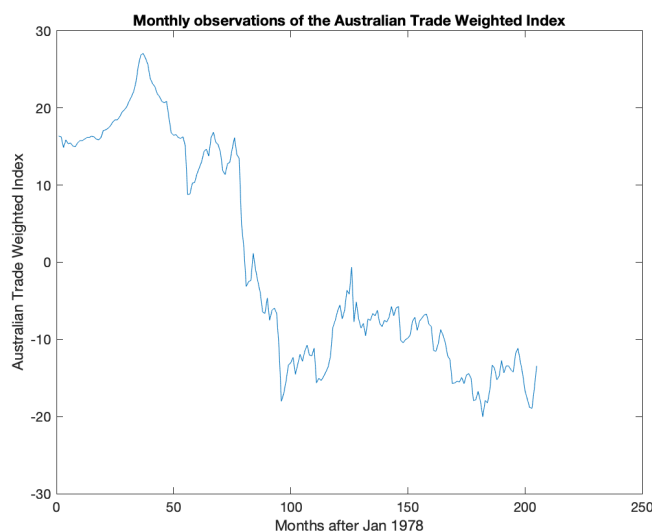

Figure 1: Monthly observations of the Australian Trade Weighted Index

In Figure 1 we we can see the the monthly observations of the Australian Trade Weighted Index. From this graph it does not appear to be a stationary process. For an lag set to say $h = 10$, if we look at the relation between $t = 25$ and $t = 75$, we see that the covariance is positive between X_{25} and X_{35} and fairly strongly negatively correlated between X_{75} and X_{85}. This shows informally that we are probably not dealing with a stationary time series as we would expect the covariance of $\gamma_X(t + h, t)$ not to depend on t.
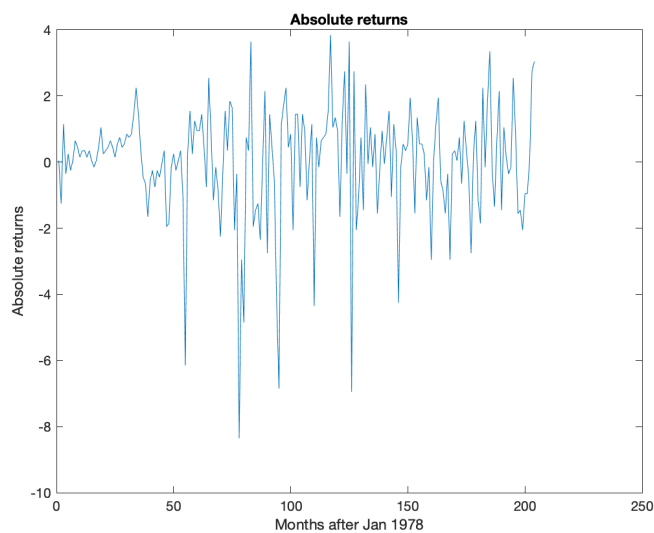
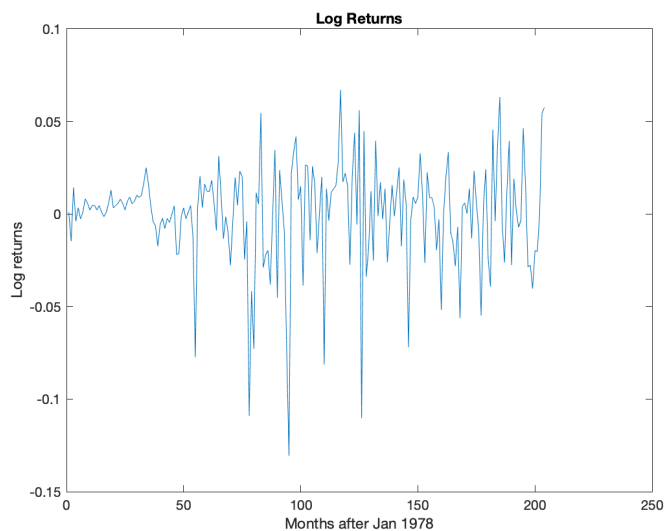Figure 2: Absolute returns $Y_t = X_{t+1} - X_t$

Figure 3: Log returns $Z_t = \log(X_{t+1}) - \log(X_t)$

For both the absolute returns in Figure 2 and the log returns in Figure 3 it appears that the processes could be stationary. There do not seem to be any pairs of points where the covariance of X_{t+h} and X_t would be dependent on

the value of t. The range of t from $0 \leq t40$ does appear slightly different from the rest of the series but not enough for us to think it would not be stationary. Overall both series actually do not look like they have very much correlation at all between which could be an indication of independent and identically distributed time series.

Problem 2

To test whether the series are iid or not we use the Ljung-Box test that compares the test statistic

$$\lambda = n(n+2) \sum_{i=1}^{h} \frac{\hat{\rho}(i)^2}{n-i}$$

Which has a χ_h^2 distribution under the null hypothesis (that the data comes from and iid distribution We want to test with $h = 20$ and $\alpha = 0.05$ so we compare λ again the critical value, $\chi_{1-\alpha,h}^2$. That is, we reject the null hypothesis if

$$\lambda > \chi_{.95,20}^2 = 31.41$$

For the standardized exchange rate data we found that $\lambda_1 = 3005$ which has a p-value of essentially 0. We can reject the null hypothesis and concluded that the exchange rate data is not iid.

For the absolute returns, we found that $\lambda_2 = 29.09$. That corresponds with a p-value of $0.086 > 0.05$. Here we fail to reject the null hypothesis (but we are fairly close to .05). There is some evidence to support the claim that the absolute returns are iid.

The log returns have a test statistic $\lambda_3 = 23.19$ which has a p-value of 0.279. Here we fail to reject the null hypothesis as well which is consistent with the log returns being iid.

The ACF are plotted in Figures 4, 5, and 6. The dashed lines represent $\pm 1.96/\sqrt{n}$. If the data was an iid sequence we would expect 95% (approximately 10 point) of the sample to fall within these bounds [1]. We see that almost all the points are within these bounds. This is consistent with out findings in the Ljung-Box test. We can see a steady decrease in the ACF for the exchange data which indicates that we could probably predict these values better than the absolute returns and log returns.
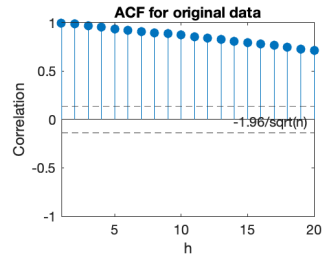
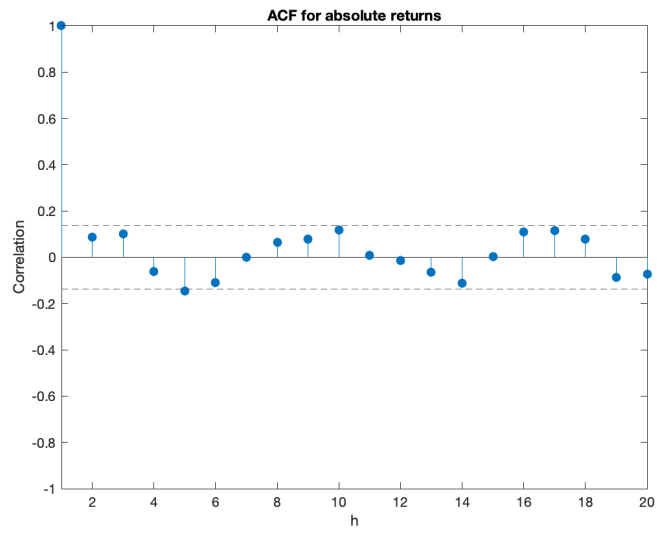Figure 4: ACF for Exchange Data $h = 0, \ldots, 20$

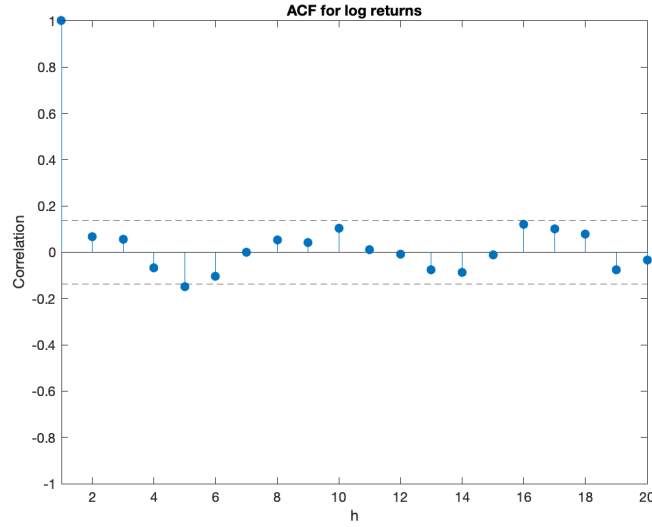Figure 5: ACF for Absolute Returns $h = 0, \ldots, 20$

Figure 6: ACF for Log Returns $h = 0, \ldots, 20$

Problem 3

We split the log returns into a training (first 102 observations) and the test set $(t = 103, \ldots, 204)$. Next we computed the autocovariance function $\hat{\gamma}$ using only the data from the training data set for each $h = 1, \ldots, 20$.

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (X_{t+h} - \overline{X})(X_t - \overline{X}) = \sum_{t=1}^{n-h} X_{t+h} X_t$$

We use these values in the sample covariance matrix

$$\hat{\Gamma}_{20} = \begin{bmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \hat{\gamma}(2) & \ldots & \hat{\gamma}(19) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \hat{\gamma}(0) & \ldots & \hat{\gamma}(18) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}(19) & \hat{\gamma}(18) & \hat{\gamma}(17) & \ldots & \hat{\gamma}(0) \end{bmatrix}$$

and solve for the coefficients, $a_1, \ldots, a_n$

$$\hat{\Gamma}_{20}(a_1, \ldots, a_n)^T = (\hat{\gamma}(20), \hat{\gamma}(19), \ldots, \hat{\gamma}(1))^T$$

Which results in the linear forecasts

$$
\begin{aligned}
b_n^l(z_{n-1}, \ldots, z_{n-20}) &= a_1 z_{n-1} + a_2 z_{n-2} + \ldots + a_{20} z_{n-20} \\
&= -0.098 z_{n-1} + -0.018 z_{n-2} + -0.179 z_{n-3} + 0.101 z_{n-4} + \\
&\quad 0.015 z_{n-5} + 0.088 z_{n-6} + 0.115 z_{n-7} + -0.097 z_{n-8} + \\
&\quad -0.036 z_{n-9} + 0.061 z_{n-10} + 0.013 z_{n-11} + 0.096 z_{n-12} + \\
&\quad 0.098 z_{n-13} + -0.048 z_{n-14} + 0.051 z_{n-15} + -0.035 z_{n-16} + \\
&\quad 0.005 z_{n-17} + -0.158 z_{n-18} + 0.062 z_{n-19} + 0.192 z_{n-20}
\end{aligned}
$$

In Figure 7 we show the predictions in red and the actual values in black. We can see as we go later in time the predictions tend towards the mean value of 0.
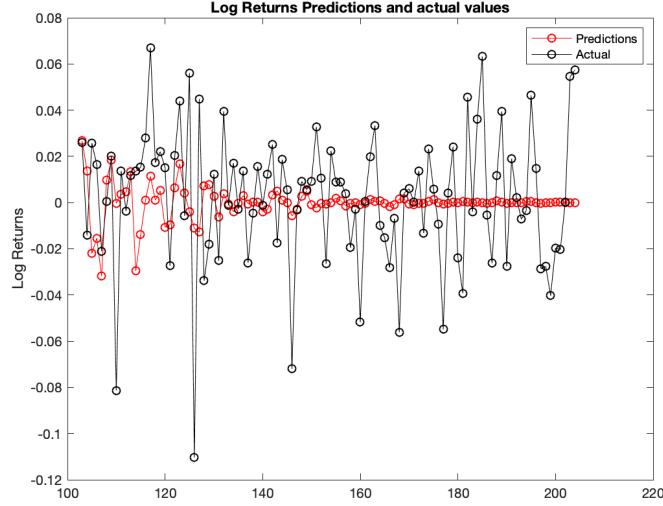

Figure 7: Linear predictions for $n = 103, \ldots, 204$

We computed the mean squared error

$$
\frac{1}{102} \sum_{n=103}^{204} (b_n^l(z_{n-1}, \ldots, z_{n-20}) - z_n)^2 = 0.00088
$$

where the prediction of just using the mean is

$$
\frac{1}{102} \sum_{103}^{204} z_n^2 = 0.00089
$$

We can see the the linear forecast MSE is almost exactly the same as the mean prediction. We showed in problem 2 that the log returns looked iid. This means

that we have no correlation (besides noise) between the points in time. We do not have any signal to predict from which results in the the linear forecasts being almost the same as the mean prediction. We can see this behavior in Figure 7 by noticiing that the predictions tend toward 0 as we go towards the end of the predictions.

Problem 4

We created a QQ-plot, Figure 8 for the log returns to check if the data was consistent with a normal distribution. We sampled points from the standard normal distribution and compared against the order statistics for the normalized log returns. We can see that the distribution of log returns does not look normal.
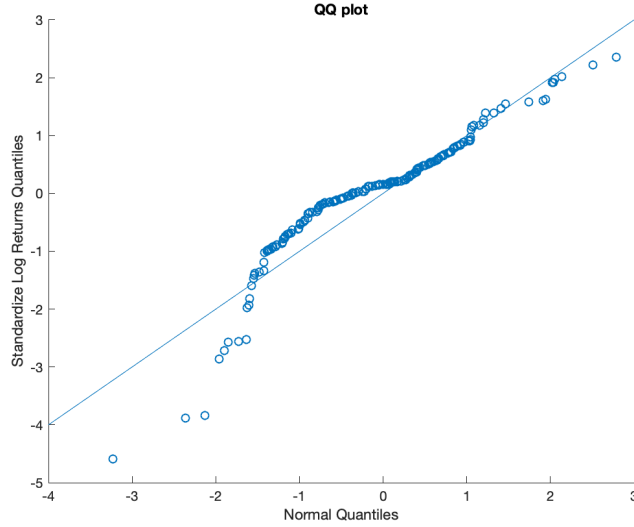

Figure 8: QQ-plot for log returns

We then tested the series

$$|Z_t| = |\log(X_{t+1}) - \log(X_t)|$$

using the Ljung-box test. We found the critical value $\lambda_4 = 76.28 > 31.41$ so we can reject the null hypothesis that the absolute value of the returns are iid. The p-value is 1.66×10^{-8}.

Problem 5

Part a

$$E[X_t] = E[\cos(\omega t + Y)]$$
$$= E[\cos(\omega t)\cos(Y) - sin(\omega t)sin(Y)] \qquad (1)$$
$$= \cos(\omega t)E[\cos(Y)] - sin(\omega t)E[sin(Y)]$$
$$= \cos(\omega t) \cdot 0 - sin(\omega t) \cdot 0 = 0 \qquad (2)$$

(1) Trignonometric identity

(2)

$$E[\cos(Y)] = \int_{-\infty}^{\infty} \cos(x)\mathbb{I}_{[-\pi,\pi]}dx$$
$$= \int_{-\pi}^{\pi} \cos(x)dx = [sin(x)]_{-\pi}^{\pi} = 0 + 0 = 0$$

$$E[\sin(Y)] = \int_{-\infty}^{\infty} \sin(x)\mathbb{I}_{[-\pi,\pi]}dx = \int_{-\pi}^{\pi} sin(x)dx$$
$$= [-cos(x)]_{-\pi}^{\pi} = -\cos(\pi) - (-cos(-\pi))$$
$$= -(-1) - (-(-1)) = 1 - 1 = 0$$

$$\gamma_X(r,s) = Cov(X_r, X_s) = E[\cos(wr + Y)cos(ws + Y)]$$
$$= E[(\cos(wr)\cos(Y) - \sin(wr)\sin(Y))(\cos(ws)\cos(Y) - \sin(ws)\sin(Y)]$$
$$= E[\cos^2(Y)\cos(wr)\cos(ws) + \cos(wr)\cos(Y)\sin(wr)\sin(Y) - \sin(wr)\sin(Y)\cos(ws)\cos(Y)$$
$$+ \sin^2(Y)sin(wr)sin(ws)]$$
$$= E[\cos^2(Y)]\cos(wr)\cos(ws) + E[\cos(Y)\sin(Y)](\cos(wr)\sin(wr) - \sin(ws)\cos(ws))$$
$$+ E[\sin^2(Y)sin(wr)sin(ws)]$$
$$\underbrace{=}_{Additional\ calculation} \pi(\cos(wr)\cos(ws) - \sin(wr)\sin(ws))$$
$$\underbrace{=}_{trig.identity} \pi\cos(w(r-s))$$
$$= \pi\cos(w((r+h) - (s+h))$$
$$= \gamma_X(r+h, s+h)$$

This shows the process to be stationary.

We can derive from the calculation of Cov that for the Autocovariance holds the following:

$\gamma_X(h) = \pi \cos(wh)$

Additional Calculation:

In the first part we show $E[\cos^2(Y)] = \pi$ and $E[\sin^2(Y)] = \pi$. In the second part we show $E[\sin(Y)\cos(Y)] = 0$

1) $Y \sim \mathcal{U}[-\pi, \pi]$

$$\int \cos^2(y)dy = \int \cos(y)\cos(y) =$$

$$\underbrace{=}_{Int.\ by\ parts} \sin(y)\cos(y) - \int -\sin(y)\sin(y)$$

$$= \cos(y)\sin(y) + \int 1 - \cos^2(y)$$

$$= \cos(y)\sin(y) + y + \int \cos^2(y)dy$$

$$\Rightarrow 2\int \cos^2(y)dy = \cos(y)\sin(y) + y$$

$$\Rightarrow \int \cos^2(y)dy = \frac{1}{2}(\cos(y)\sin(y) + y)$$

$$\Rightarrow E[\cos^2(Y)] = \int_{-\pi}^{\pi} \cos^2(y)dy = [\frac{1}{2}(\cos(y)\sin(y) + y)]_{-\pi}^{\pi} = \pi$$

$$E[\sin^2(Y)] = \cdots = [\frac{1}{2}(-\cos(y)\sin(y) + y)]_{-\pi}^{\pi} = \pi$$

2) $Y \sim \mathcal{U}[-\pi, \pi]$

$$\int \sin(y)\cos(y)dy \underbrace{=}_{u=\sin(y),\ du=cos(y)dy} \int u du$$

$$= \frac{1}{2}u^2 + C = \sin^2(y) + C$$

$$\Rightarrow E[\sin(Y)\cos(Y)] = [\sin^2(y)]_{-\pi}^{\pi} = 0 - 0 = 0$$

References

[1] Peter J. Brockwell and Richard A. Davis. *Introduction to Time Series and Forecasting.* Springer, 3rd edition, 2016.