

Time Series Project 1

Stefan Eng and Franz Hartleitner

April 2019

Problem 1

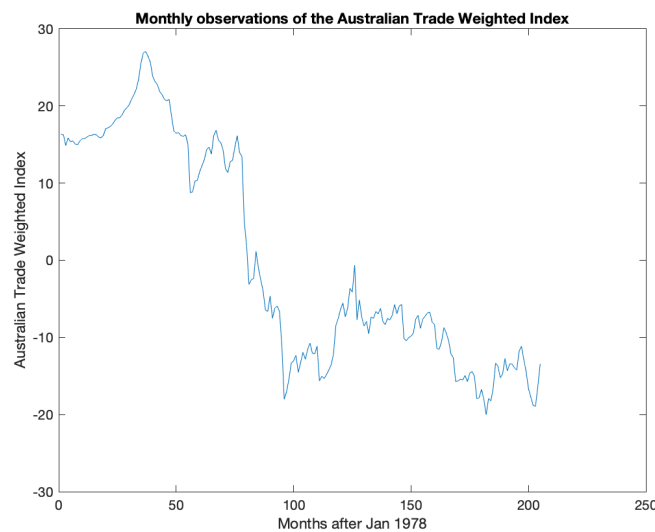


Figure 1: Monthly observations of the Australian Trade Weighted Index

In Figure 1 we can see 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 .

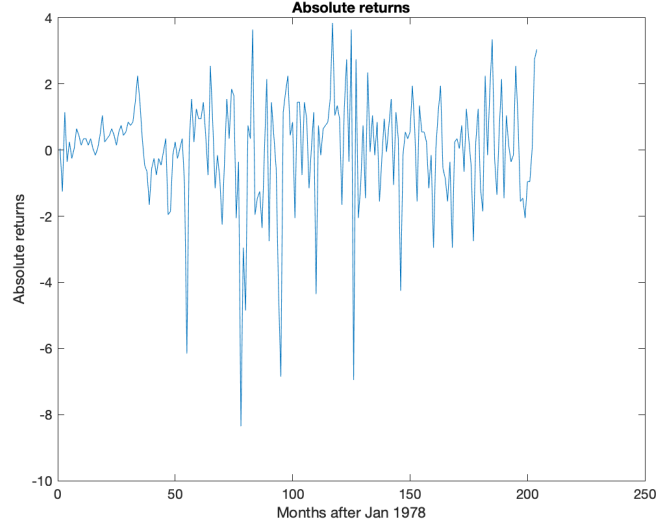


Figure 2: Absolute returns $Y_t = X_{t+1} - X_t$

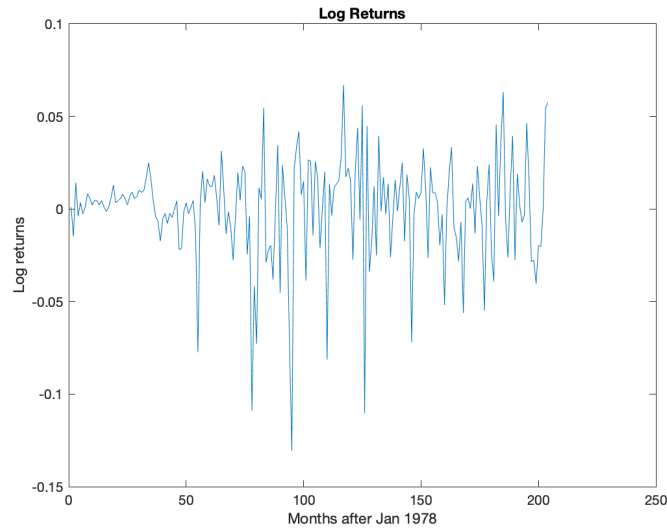


Figure 3: Log returns $Z_t = \log(X_{t+1}) - \log(X_t)$

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 $0 \leq t < 40$ 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 an iid distribution). We want to test with $h = 20$ and $\alpha = 0.05$ so we compare λ against 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 conclude 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 our 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.

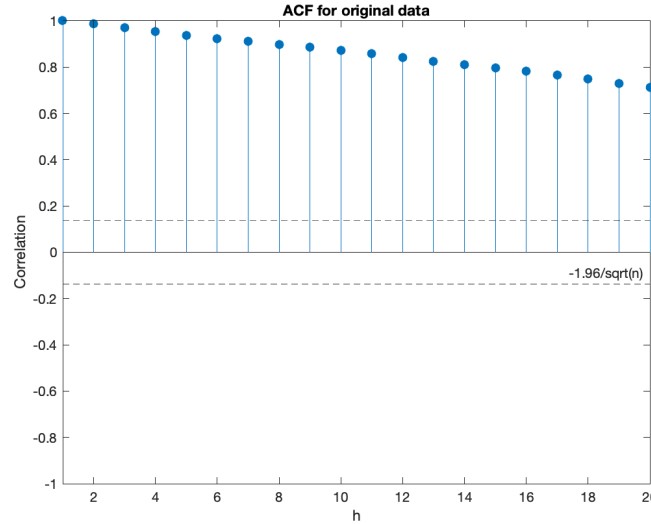


Figure 4: ACF for Exchange Data $h = 0, \dots, 20$

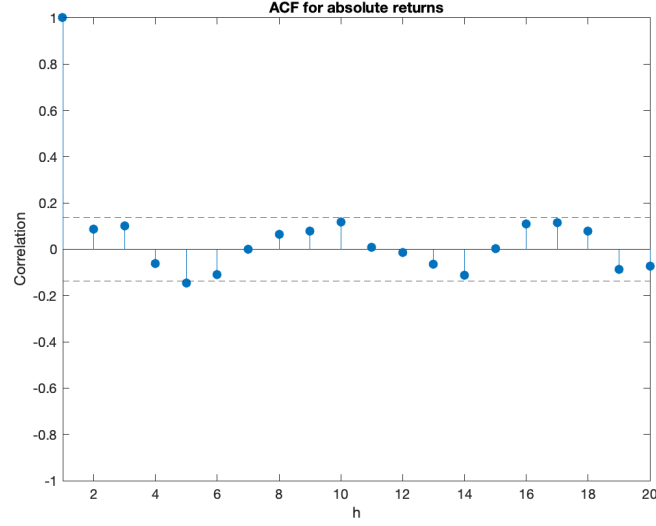


Figure 5: ACF for Absolute Returns $h = 0, \dots, 20$

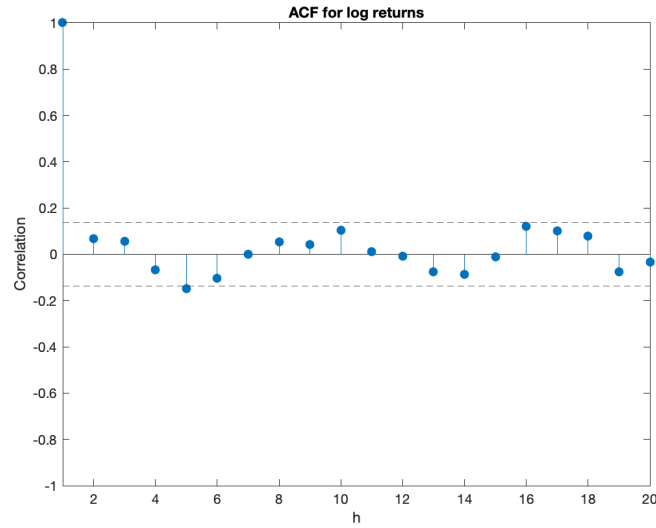


Figure 6: ACF for Log Returns $h = 0, \dots, 20$

Problem 3

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

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

We use these values in the sample covariance matrix

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

and solve for the coefficients, a_1, \dots, a_n

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

Which results in the linear forecasts

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

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.

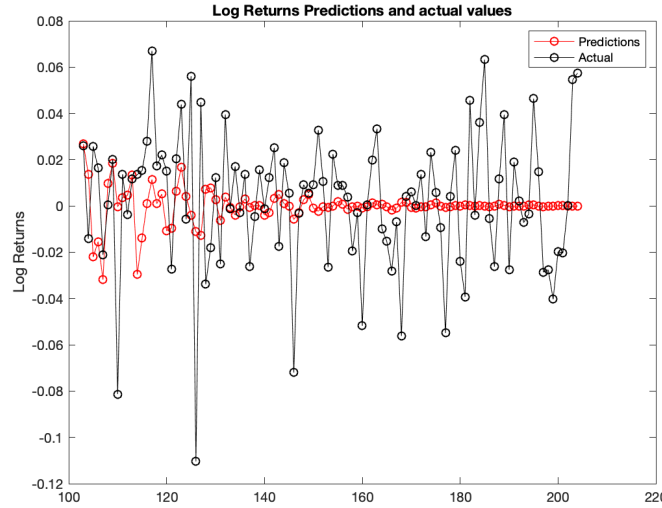


Figure 7: Linear predictions for $n = 103, \dots, 204$

We computed the mean squared error

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

where the prediction of just using the mean is

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

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 noticing 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. We confirmed this result by using Matlab's `kstest` which rejected the null hypothesis that the normalized log returns come from a standard normal distribution.

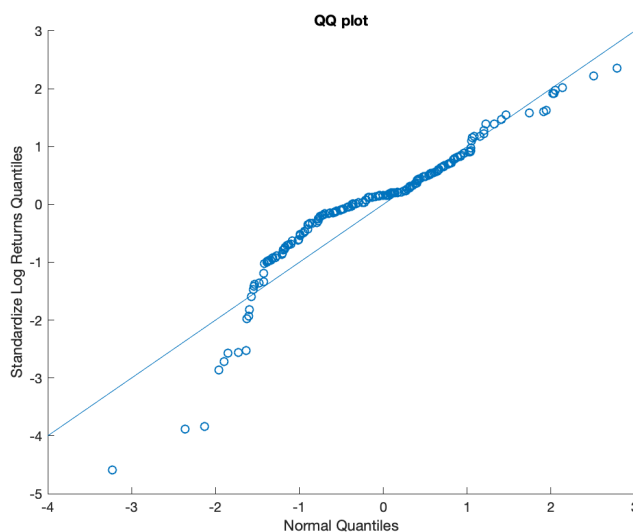


Figure 8: QQ-plot for log returns

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

$$\begin{aligned} E[X_t] &= E[\cos(\omega t + Y)] \\ &= E[\cos(\omega t) \cos(Y) - \sin(\omega t) \sin(Y)] \end{aligned} \tag{1}$$

$$\begin{aligned} &= \cos(\omega t) E[\cos(Y)] - \sin(\omega t) E[\sin(Y)] \\ &= \cos(\omega t) \cdot 0 - \sin(\omega t) \cdot 0 = 0 \end{aligned} \tag{2}$$

(1) Sum-Difference Formula

(2)

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

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

$$\begin{aligned}
\gamma_X(t, t+h) &= Cov(X_t, X_{t+h}) = E[\cos(\omega t + Y) \cos(\omega(t+h) + Y)] \\
&= E[(\cos(\omega t) \cos(Y) - \sin(\omega t) \sin(Y))(\cos(\omega t + \omega h) \cos(Y) - \sin(\omega t + \omega h) \sin(Y))] \\
&= E[\cos^2(Y) \cos(\omega t) \cos(\omega t + \omega h) + \sin^2(Y) \sin(\omega t) \sin(\omega t + \omega h) \\
&\quad + \cos(\omega t) \sin(\omega t + \omega h) \sin(Y) \cos(Y) - \sin(\omega t) \cos(\omega t + \omega h) \sin(Y) \cos(Y)] \\
&= \underbrace{E[\cos^2(Y)]}_{\pi \text{ by Appendix A.1}} \cos(\omega t) \cos(\omega t + \omega h) + \underbrace{E[\sin^2(Y)]}_{\pi \text{ by Appendix A.1}} \sin(\omega t) \sin(\omega t + \omega h) + \\
&\quad \underbrace{E[\sin(Y) \cos(Y)]}_{0 \text{ by Appendix A.2}} [\cos(\omega t) \sin(\omega t + \omega h) - \sin(\omega t) \cos(\omega t + \omega h)] \\
&= \pi [\cos(\omega t) \cos(\omega t + \omega h) + \sin(\omega t) \sin(\omega t + \omega h)] \\
&= \pi [\cos(\omega t) [\cos(\omega t) \cos(\omega h) - \sin(\omega t) \sin(\omega h)] + \\
&\quad \sin(\omega t) [\sin(\omega t) \cos(\omega h) + \sin(\omega h) \cos(\omega t)]] \\
&= \pi [\cos^2(\omega t) \cos(\omega h) + \sin^2(\omega t) \cos(\omega h) + \\
&\quad \sin(\omega h) \cos(\omega t) \sin(\omega t) - \sin(\omega h) \cos(\omega t) \sin(\omega t)] \\
&= \pi \cos(\omega h) [\cos^2(\omega t) + \sin^2(\omega t)] \\
&= \pi \cos(\omega h)
\end{aligned}$$

since $\gamma_X(t, t+h)$ which does not depend on t . It follows that the autocovariance function for X_t is $\gamma_X(h) = \pi \cos(\omega h)$.

Finally, We can also see that $\text{Var}(\cos(\omega t + Y)) < +\infty$

$$\begin{aligned}
\text{Var}(\cos(\omega t + Y)) &= \gamma(0) \\
&= \pi \cos(0) = \pi < \infty
\end{aligned}$$

This completes the proof that the process to be stationary.

Part b

We want to solve

$$\begin{aligned}
\begin{bmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} &= \begin{bmatrix} \gamma(1) \\ \gamma(2) \end{bmatrix} \\
\begin{bmatrix} \pi & \pi \cos(\omega) \\ \pi \cos(\omega) & \pi \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} &= \begin{bmatrix} \pi \cos(\omega) \\ \pi \cos(2\omega) \end{bmatrix}
\end{aligned}$$

$$\pi(a_1 + \cos(\omega)a_2) = \pi \cos(\omega) \implies a_1 = \cos(\omega)(1 - a_2) \quad (3)$$

$$\pi(\cos(\omega)a_1 + a_2) = \pi \cos(2\omega) \quad (4)$$

Plugging in (3) into (4) we get

$$\begin{aligned}
\cos^2(\omega) - \cos^2(\omega)a_2 + a_2 &= \cos(2\omega) \\
\cos^2(\omega) + (1 - \cos^2(\omega))a_2 &= \cos(2\omega) \\
a_2 &= \frac{\cos(2\omega) - \cos^2(\omega)}{1 - \cos^2(\omega)} \\
&= \frac{2\cos^2(\omega) - 1 - \cos^2(\omega)}{1 - \cos^2(\omega)} \\
&= -1
\end{aligned} \tag{5}$$

Line (5) follows because $\cos(2\omega) = 2\cos^2(\omega) - 1$. It then follows that $a_1 = 2\cos(\omega)$. Thus,

$$b_{n+1}^l(X_{n-1}, X_n) = 2X_n \cos(\omega) - X_{n-1}$$

Part c

The mean squared error is

$$\begin{aligned}
E[(X_{n+1} - b_{n+1}^l(X_{n-1}, X_n))^2] &= E[(X_{n+1} - 2X_n \cos(\omega) + X_{n-1})^2] \\
&= E[(X_{n+1})^2 - 2X_n X_{n+1} \cos(\omega) + X_{n-1} X_{n+1} \\
&\quad - 2X_n X_{n+1} \cos(\omega) + 4X_n^2 \cos^2(\omega) \\
&\quad - 2X_{n-1} X_n \cos(\omega) + X_{n-1} X_{n+1} \\
&\quad - 2X_{n-1} X_n \cos(\omega) + X_{n-1}^2] \\
&= E[X_{n+1}^2] - 4E[X_n X_{n+1}] \cos(\omega) + 2E[X_{n-1} X_{n+1}] \\
&\quad - 4E[X_{n-1} X_n] \cos(\omega) + 4E[X_n^2] \cos^2(\omega) + E[X_{n-1}^2] \\
&= \pi - 4\gamma(1) \cos(\omega) + 2\gamma(2) \\
&\quad - 4\gamma(1) \cos(\omega) + 4\pi \cos^2(\omega) + \pi \\
&= 2\pi - 8\pi \cos^2(\omega) + 4\pi \cos^2(\omega) + 2\pi \cos(2\omega) \\
&= 2\pi - 4\pi \cos^2(\omega) + 2\pi(2\cos^2(\omega) - 1) \\
&\quad \text{(Double Angle Formula)} \\
&= 0
\end{aligned}$$

Appendices

A Additional Calculation

A.1 $E[\cos^2(Y)] = \pi$ and $E[\sin^2(Y)] = \pi$

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

$$\begin{aligned}
E[\cos^2(Y)] &= \int_{-\pi}^{\pi} \cos^2(y) dy \\
&= \int_{-\pi}^{\pi} \frac{1 + \cos(2y)}{2} dy && \text{half angle formula} \\
&= \frac{1}{2}y + \frac{1}{4}\sin(2y) \Big|_{-\pi}^{\pi} \\
&= \frac{1}{2}\pi - (-\frac{1}{2}\pi) = \pi
\end{aligned}$$

$$\begin{aligned}
E[\sin^2(Y)] &= \int_{-\pi}^{\pi} \cos^2(y) dy \\
&= \int_{-\pi}^{\pi} \frac{1 + \cos(2y)}{2} dy && \text{half angle formula} \\
&= \frac{1}{2}y + \frac{1}{4}\sin(2y) \Big|_{-\pi}^{\pi} \\
&= \pi
\end{aligned}$$

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

A.2 $E[\sin(Y) \cos(Y)] = 0$

$$\begin{aligned}
\int \sin(y) \cos(y) dy &\stackrel{u=\sin(y), \ du=\cos(y)dy}{=} \int u du \\
&= \frac{1}{2}u^2 + C = \frac{1}{2}\sin^2(y) + C
\end{aligned}$$

$$\Rightarrow E[\sin(Y) \cos(Y)] = \left[\frac{1}{2}\sin^2(y) \right]_{-\pi}^{\pi} = 0 - 0 = 0$$

A.3 project1.m

```

data = importdata('exchangerate.mat');
n_data = length(data);

% Problem 1
x_t = data(1:end - 1);
x_t_1 = data(2:end);

% Compute absolute and log returns
abs_returns = x_t_1 - x_t;
log_returns = log(x_t_1) - log(x_t);

n_returns = length(abs_returns);

corrected_abs_returns = abs_returns - mean(abs_returns);
corrected_log_returns = log_returns - mean(log_returns);
corrected_data = data - mean(data);

% Plot exchange rate data
figure;
plot(corrected_data);
ylabel("Australian Trade Weighted Index");
xlabel("Months after Jan 1978");
title("Monthly observations of the Australian Trade Weighted Index");
saveas(gcf, 'plots/exchangedata.png');
figure;

% Plot absolute returns
plot(corrected_abs_returns);
xlabel("Months after Jan 1978");

```

```

ylabel("Absolute_returns");
title("Absolute_returns");
saveas(gcf, 'plots/abs_returns.png');
figure;

% Plot log returns
plot(corrected_log_returns);
xlabel("Months_after_Jan_1978");
ylabel("Log_returns");
title("Log_Returns");
saveas(gcf, 'plots/log_returns.png');

% Problem 2
crit_val = chi2inv(0.95, 20);
[lambda_data, acf_data] = ljungbox(corrected_data, 20);
[lambda_abs, acf_abs] = ljungbox(corrected_abs_returns, 20);
[lambda_log, acf_log] = ljungbox(corrected_log_returns, 20);

% Hypothesis test
chi2cdf(lambda_data, 20, 'upper');
chi2cdf(lambda_abs, 20, 'upper');
chi2cdf(lambda_log, 20, 'upper');

% IID should be 0 covariance
figure;
stem(acf_data, 'filled');
% Draw  $\pm 1.96 / \sqrt{n}$  lines
yline(-1.96 / sqrt(n_data), '—', '-1.96/sqrt(n)');
yline(1.96 / sqrt(n_data), '—');
title("ACF_for_original_data");
ylabel("Correlation");
xlabel("h");
axis([1 20 -1 1]);
saveas(gcf, 'plots/acf_exchange.png');

% ACF Abs returns plot
figure;
stem(acf_abs, 'filled');
% Draw  $\pm 1.96 / \sqrt{n}$  lines
yline(1.96 / sqrt(n_returns), '—');
yline(-1.96 / sqrt(n_returns), '—');
title("ACF_for_absolute_returns");
ylabel("Correlation");
xlabel("h");
axis([1 20 -1 1]);
saveas(gcf, 'plots/acf_abs_returns.png');

% ACF Log Returns
figure;
stem(acf_log, 'filled');
% Draw  $\pm 1.96 / \sqrt{n}$  lines
yline(1.96 / sqrt(n_returns), '—');
yline(-1.96 / sqrt(n_returns), '—');
title("ACF_for_log_returns");
ylabel("Correlation");
xlabel("h");

```

```

axis([1 20 -1 1]);
saveas(gcf, 'plots/acf-log-returns.png');

% Problem 3
training = corrected_log_returns(1:102);
test = corrected_log_returns(103:end);
[gm, train-gamma-mat, train-acf] = acvf(training, 20);

% Compute coefficients and make predictions
% First coefficient, a_0 is zero since mean is zero.
coefs = train-gamma-mat \ flip(gm(2:end)');

preds = zeros(n_returns, 1);
preds(1:102) = training;
for i = 103:n_returns
    preds(i) = preds(i-1:-1:i-20)' * coefs;
end

% Print out the coefficients nicely for latex
coef_format = arrayfun(@(x) sprintf("%.3f", x), coefs);
zs_format = arrayfun(@(x) sprintf("z_{n-%d}", x), 1:20);
blp_format = strcat(coef_format', zs_format);
% add plus between coefficients
blp_format = strjoin(blp_format, ' + ');

% Plot the predictions (red) and the actual values (black)
figure;
preds_plot = plot(103:n_returns, preds(103:n_returns), 'o');
preds_plot.Color = "red";
hold on;
actual_plot = plot(103:n_returns, test, 'o');
actual_plot.Color = "black";
title("Log_Returns_Predictions_and_actual_values");
ylabel("Log_Returns");
legend("Predictions", "Actual");
saveas(gcf, 'plots/log-returns-preds.png');

% Plot the residuals
% figure;
% stem(103:n_returns, preds(103:n_returns) - test, 'filled');
% title("Predicted - Actual (Residuals)");

% mean squared error
forecast_mse = mean((preds(103:n_returns) - test).^2);
mean_mse = mean(test.^2);

% Not much difference in the forecasted MSE and the naive mean prediction
% Since log returns look IID it makes sense that mean prediction is almost
% as good

% Problem 4
% qqplot(corrected_log_returns);
% Standardized log return (divided off standard deviation)
log_returns_std = sort(corrected_log_returns / std(corrected_log_returns));
% Generate samples from normal distribution
norm_sample = sort(normrnd(0, 1, [1, n_returns]));

```

```

% Plot QQ-plot
figure;
scatter(norm_sample, log_returns_std);
hold on;
refline(1,0);
title("QQ-plot");
ylabel("Standardize_Log_Returns_Quantiles");
xlabel("Normal_Quantiles");
saveas(gcf, 'plots/qqplot.png');

% Test of Normality by Kolmogorov Smirnov Test:
% h = kstest(x) returns a test decision for the null hypothesis that the
% data in vector x comes from a standard normal distribution, against
% the alternative that it does not come from such a distribution, using
% the one-sample Kolmogorov-Smirnov test. The result h is 1 if the test
% rejects the null hypothesis at the 5 significance level, or 0 otherwise.
[h,p] = kstest(corrected_log_returns);

% Second part, repeat problem 2 but for |Z|
log_abs_returns = abs(log(x_t_1) - log(x_t));
log_abs_returns = log_abs_returns - mean(log_abs_returns);
[lambda_log_abs, ~] = ljungbox(log_abs_returns, 20);
log_abs_pval = chi2cdf(lambda_log_abs, 20, 'upper');

```

A.4 acvf.m

```

function [gm, gamma_mat, acf] = acvf(data, h)
% gamma_mat is sample covariance matrix
% gm is the gamma values
gm = zeros(1,h+1);
n = length(data);
mx = mean(data);

for i = 0:h
    gm(i + 1) = (data(1 + i:end) - mx)' * (data(1:end - i) - mx) / n;
end

gamma_mat = zeros(h, h);
for i = 1:h
    for j = i:h
        res = gm((j - i) + 1);
        gamma_mat(i,j) = res;
        gamma_mat(j,i) = res;
    end
end

acf = gm / gm(1);
end

```

A.5 ljungbox.m

```

function [gm, gamma_mat, acf] = acvf(data, h)
% gamma_mat is sample covariance matrix
% gm is the gamma values
gm = zeros(1,h+1);

```

```

n = length(data);
mx = mean(data);

for i = 0:h
    gm(i + 1) = (data(1 + i:end) - mx)' * (data(1:end - i) - mx) / n;
end

gamma_mat = zeros(h, h);
for i = 1:h
    for j = i:h
        res = gm((j - i) + 1);
        gamma_mat(i, j) = res;
        gamma_mat(j, i) = res;
    end
end

acf = gm / gm(1);
end

```

References

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