Machine Learning for Finance: Assignment One

Due on Tuesday, May 24, 2018

Adrian Stefan & Bradley James Kester

Contents

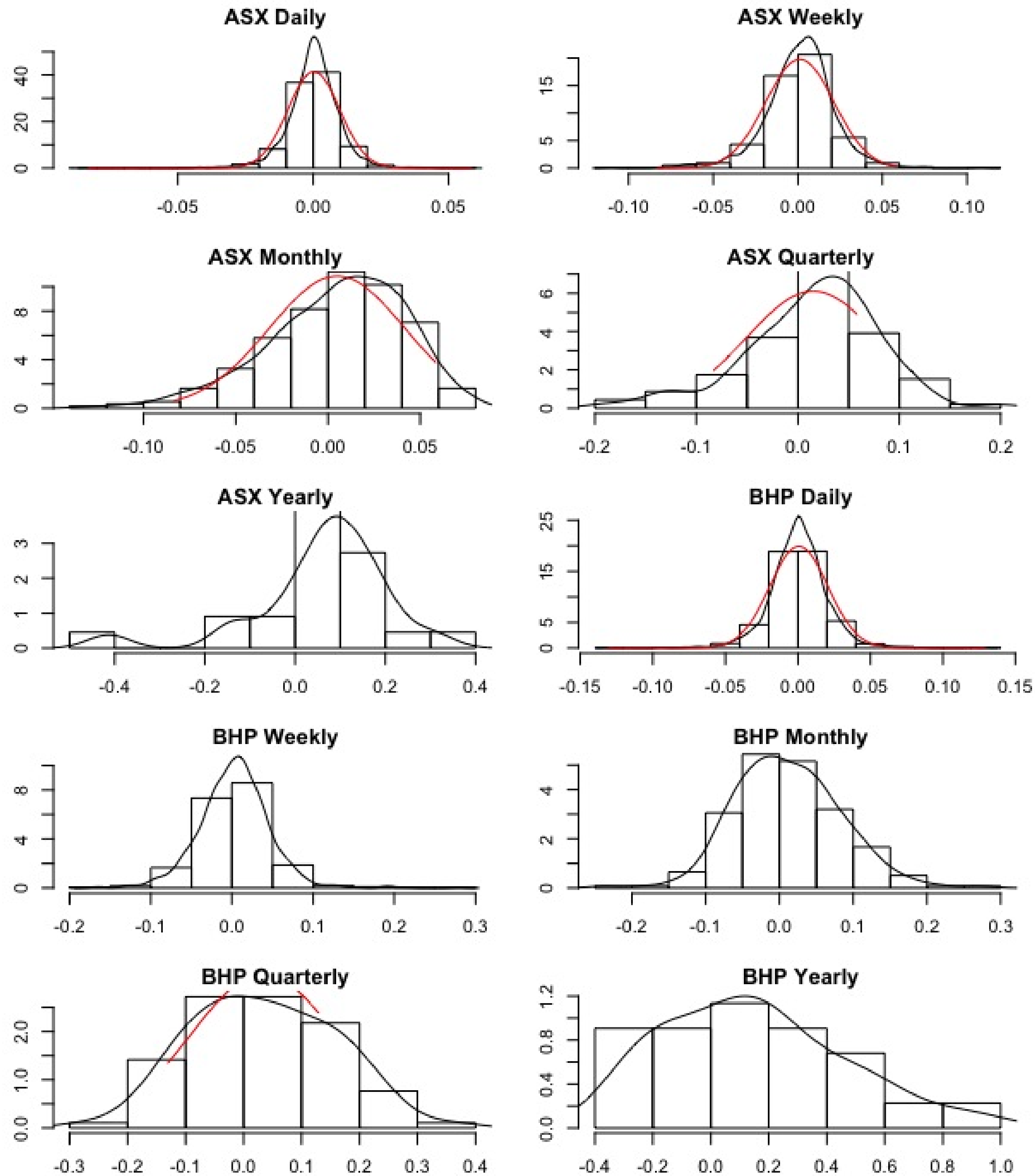
Problem 1

In this analysis we will examine the Australian Securities Exchange ASX200 index and the historical share price of BHP Billiton from the first of January, 1995 until the first of January 2018. We have elected to examine the ASX200 index instead of the ASX50 index as data for there latter commences in 2013. Summary statistics are exhibited in the tables below. Examining the ASX200 we observe that minimum and

maximum values are in increasing in absolute terms with a decreasing frequency. This is consistent for means and standard deviations. Furthermore, we observe that when examine lower frequencies the distribution of returns is increasingly negatively skewed meanwhile kurtosis is found to be decreasing. While examining aggregate normality we find across all frequencies and tests the observed p-value stays below 5% allowing us to confirm the normality hypothesis. Moving forth with BHP Billiton we find largely similar results.

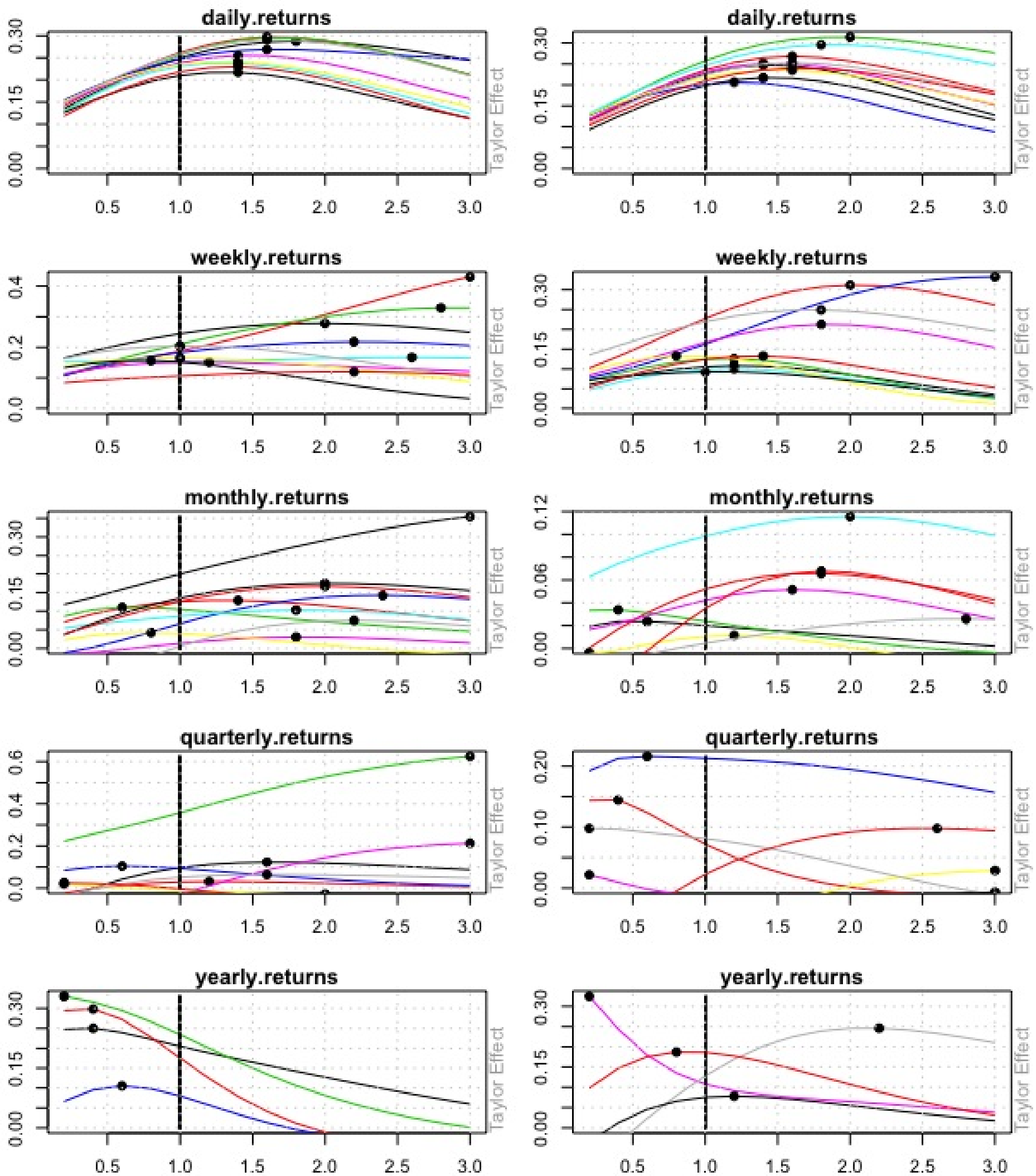
However, contrary to that observed in the ASX200 index we find that skewness is increasingly positive at higher frequencies and at frequencies of monthly and longer we do reject aggregate normality with all test with the exception of the monthly Jarque-Bera test exhibiting 5 values of greater than 5%.

Problem 1 continued on next page...


ASX Daily
0 20 40
-0.05 0.00 0.05
ASX Weekly
0 5 15
-0.10 -0.05 0.00 0.05 0.10
ASX Monthly
0 4 8
-0.10 -0.05 0.00 0.05
ASX Quarterly
0 2 4 6
-0.2 -0.1 0.0 0.1 0.2
ASX Yearly
0 1 2 3
-0.4 -0.2 0.0 0.2 0.4
BHP Daily
0 5 15 25
-0.15 -0.10 -0.05 0.00 0.05 0.10 0.15
BHP Weekly
0 4 8
-0.2 -0.1 0.0 0.1 0.2 0.3
BHP Monthly
0 2 4
-0.2 -0.1 0.0 0.1 0.2 0.3
BHP Quarterly
0.0 1.0 2.0
-0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4
BHP Yearly
0.0 0.4 0.8 1.2
-0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0

Problem 2

To empirically test the Taylor Effect we implemented the teffectplot function in R. Overall, we find that the Taylor Effect hypothesis is empirically rejected across all observed time frequencies for both the ASX200 and BHP Billiton. Although, this said we also observed that at higher frequencies the hypothesis is followed more closely than across longer time-horizons. The plot below exhibits the two time-series with the ASX200 on the left and BHP Billiton on the right.



Problem 3

In this exercise we will firstly provide a proof of the recursive notation of the EWMA volatility model. Secondly, we will calculate the EWMA for the ASX200 and BHP Billiton.

Proof

In this proof we use the following notation:

$$\sigma^2_{EWMA} = \sigma^2_\epsilon$$

The initial equation is given by:

$$\sigma^2_\epsilon = (1-\lambda)\sum_{k=1}^{\infty}\lambda^{k-1}r^2_{t-1}$$

We want to show that:

$$\sigma^2 = \lambda\sigma^2_{t-1} + (1-\lambda)r^2_{t-1}$$

First we define the volatility in the previous period:

$$\sigma^2_{t-2} = \lambda^2[\lambda\sigma^2_{t-3} + (1-\lambda)r^2_{t-3}] + \lambda(1-\lambda)r^2_{t-2} + (1-\lambda)r^2_{t-k}$$

$$= \lambda^3\sigma^2_{t-3} + \lambda^2(1-\lambda)r^2_{t-3} + \lambda(1-\lambda)r^2_{t-3} + \lambda(1-\lambda)r^2_{t-2} + (1-\lambda)r^2_{t-1}$$

$$= ... = (1-\lambda)\sum_{k=1}^{m}\lambda^{k-1}r^2_{t-k}$$

Empirics

Now we shall exhibit EWMA standard deviation in comparison with the regular standard deviation previous determined in the first problem.

Table 1: ASX200 Standard Deviations

	ASX Std	ASX EMA Std
Daily	0.01	0.02
Weekly	0.02	0.03
Monthly	0.04	0.07
Quarterly	0.07	0.12
Yearly	0.15	0.25

Therefore, we can see the EWMA model exhibits greater standar deviations for both the ASX200 and BHP Billiton across all time periods.

Table 2: BHP Billiton Standard Deviations

	BHP Std	BHP: EMA Std
Daily	0.02	0.03
Weekly	0.04	0.06
Monthly	0.07	0.11
Quarterly	0.13	0.18
Yearly	0.31	0.38

Problem 4

Firstly, note by Gauss-Marcov Theurom we know that The OLS Estimator is the best, linear unbiased estimator and therefore true and efficient when the following assumptions are satisfied:

- Error term is independently distributed and not correlated
- Linearity in parameters
- Homoskedasticity
- Zero conditional mean
- No perfect collinearity

Moving forward, we shall now formally prove why conditional expectation is the optimal forecast under squared loss and additionally, prove the conditional expectation.

The first proof we'll demonstrate is that which shows the best linear estimator for $X = X_{t+h}$ given Z is βZ^t, where $\beta = E(XZ^t)E(Z^tZ)^{-1}$.

Proof. Let $X = f(Z) + \epsilon = \beta Z^t + \epsilon$, where $\epsilon \ N(0, \sigma_\epsilon^2)$

We want to minimise the square loss:

$$\min_\beta E[(X - \beta Z^t)^2]$$

Now we must take the FOC with respect to β to obtain the optimal estimator

$$E[\Delta_\beta (X - \beta Z^t)^2] = 0$$
$$E[-2(X - \beta Z^t)Z^t] = 0$$
$$E(XZ^t) - \beta E(Z^tZ) = 0$$
$$\beta = E(XZ^t)E(Z^tZ)^{-1}$$

Problem 4 continued on next page...

There we have demonstrated the optimal β. We shall now show the optimal forecast.

Square loss:

$$
\begin{aligned}
E(X_{t+h} = f(Z))^2 = E(X_{t+h} - E(X_{t+h}|Z)E(X_{t+h}|Z) - f(Z))^2 \\
= E(X_{t+h} - E(X_{t+h}|Z))^2 + E(E(X_{t+h}|Z) - f(Z))^2 \quad + \\
2E([X_{t+h} - E(X_{t+h}|Z)][E(X_{t+h}|Z - f(Z)]
\end{aligned}
\tag{1}
$$

According to The Law of Iterated Expectation (LIE) the final term will be 0. Hence, we now left with the first and second terms of equation (1), where only the second part depends upon $f(Z)$. Thus, if $f(Z) = E(Y_{t+h}|Z)$ and subsequently $f(Z)$ equals the conditional expectation, the second term of (1) will be equal to 0 and the square loss will be minimised.

In conclusion, we have demonstrated that $E[X|Z] = Z^t\beta$:

$$
\begin{aligned}
E[X|Z] &= E[Z^T\hat{\beta}|Z] \\
&= Z^T E[\hat{\beta}|Z] \\
&= Z^T E[(Z^T Z)^{-1}(Z^T X)|Z] \\
&= Z^T E[Z^T Z]^{-1} + E[(Z^T Z)^{-1}(Z^T u|Z] \\
&= Z^T \beta
\end{aligned}
$$

Take note we have rearranged β in error form to show the proof. Furthermore, we utilised the exogeneity assumption at the beginning and finally, we relied on the conditional expectation being linear on Z.