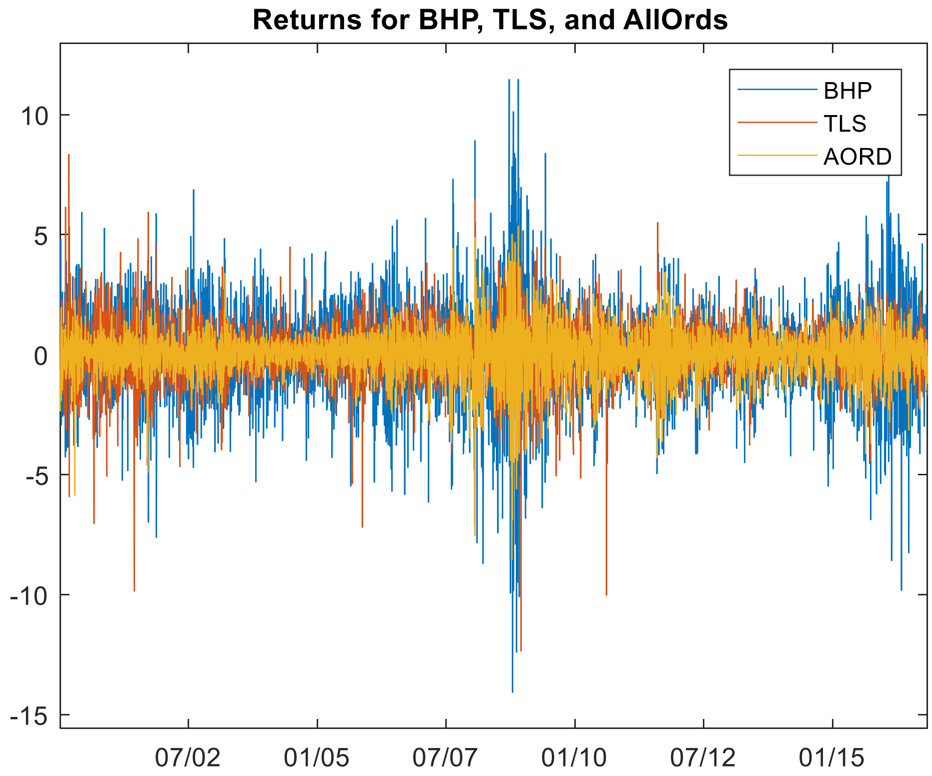
QBUS6830 Financial Time Series and Forecasting
S1, 2019

Solutions to Lab Sheet 7

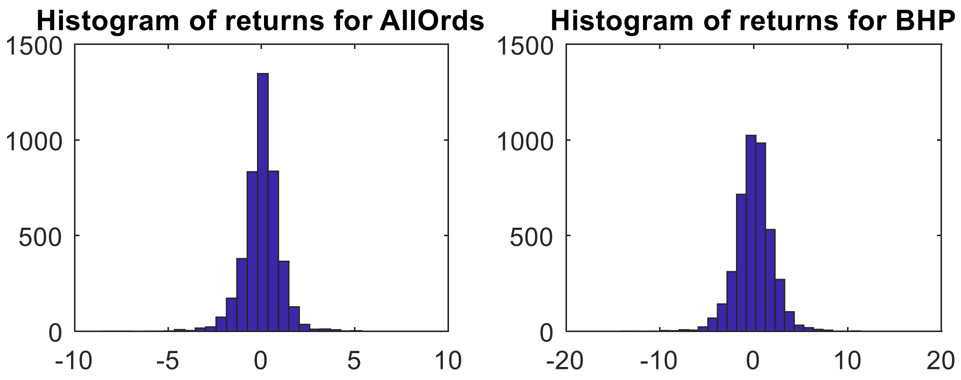
Q1 (Forecasting)
In this lab session will use daily data on BHP and Telstra returns, from Jan, 2000 to April, 2017, together with the ASX All Ordinaries index returns over the same period. The data can be found in the files BHP00-17.csv, TLS00-17.csv and AORD00-17.csv.

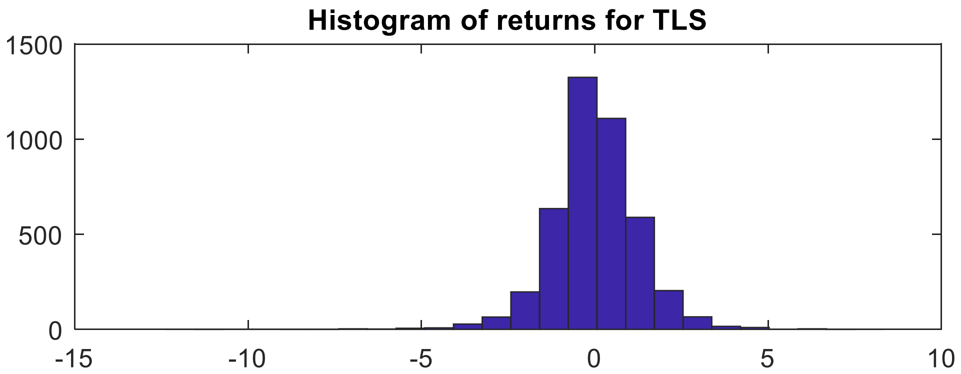
(a) Conduct a brief EDA on your assets and the market return series.

The 3 times series are plotted here:



The low volatility period leading up to 2008 is readily apparent, as is the increase in volatility marking the GFC period. The returns are clearly stationary in mean.





Clearly these assets are quite fat-tailed and each has many outlying return observations.

	AORD	BHP	TLS
Mean	0.0122	0.0397	0.0235
Median	0.0484	0.0544	0.0000
Std	0.9757	1.9598	1.2668
Min	-8.5536	-14.0772	-12.3546
Max	5.3601	11.4645	8.3427
Skew	-0.5770	-0.1760	-0.5062
Kurt	8.9446	6.5566	9.2237

All have positive means, though BHP and TLS has 0 median returns during this time period. All have negative skew estimates, no doubt influenced by having a few more large negative outlying returns, than positive. From the histograms, all appear close to symmetric. They are all clearly more fat-tailed than a Gaussian (kurtosis > 3).

(b) Set your in-sample or learning period to be the first 75% of data points in your sample. Set your forecast period to be the last 25% of your sample period. For the two asset return series, use the following methods to choose and estimate suitable models using the in-sample data only:

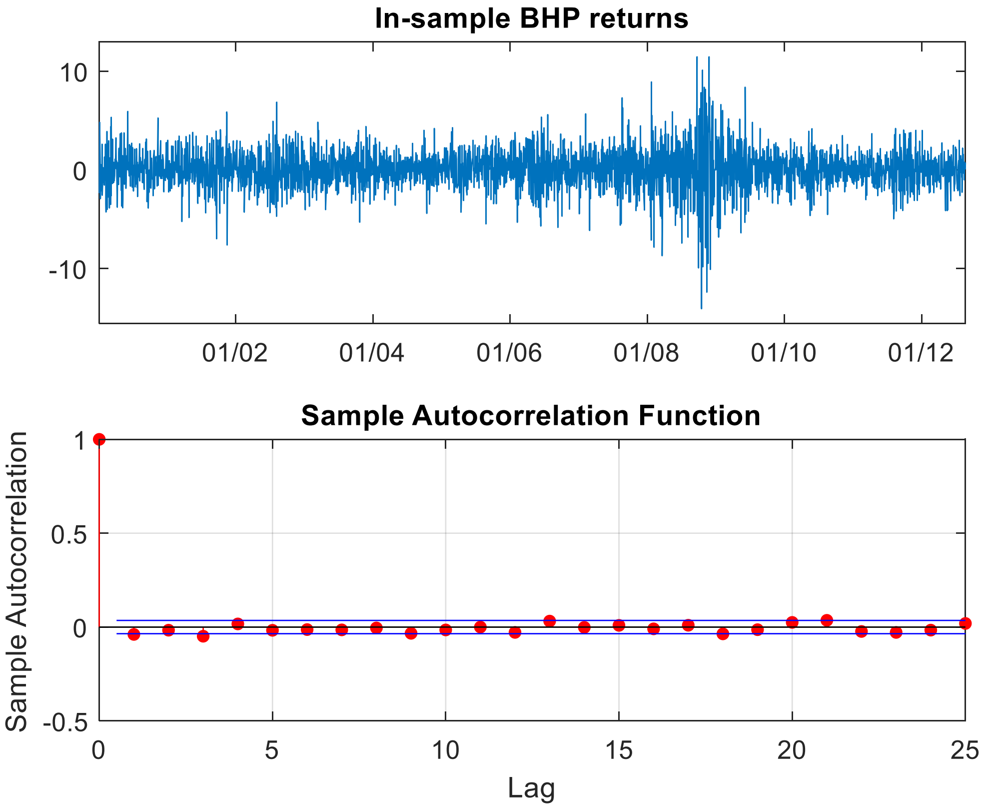
1. Naive.
2. 1 month mean for that asset.
3. 1 year mean for that asset.
4. A regression using lag 1 market return.
5. A suitably chosen ARMA model.
6. A suitably chosen Reg-ARMA model, using the lag 1 market index return series.
7. A Reg-AR(1) plus 1 suitable seasonal lag of the asset return series; e.g. for daily data a seasonal lag would be 5 (the same day last week).

<u>For BHP</u>

The regression model is estimated as:

$BHP_t = 0.0572 - 0.1019\ AORD_{t-1} + \varepsilon_t$; $R^2 = 0.27\%$, $SER = 2.01\%$

An ACF plot for BHP returns shows

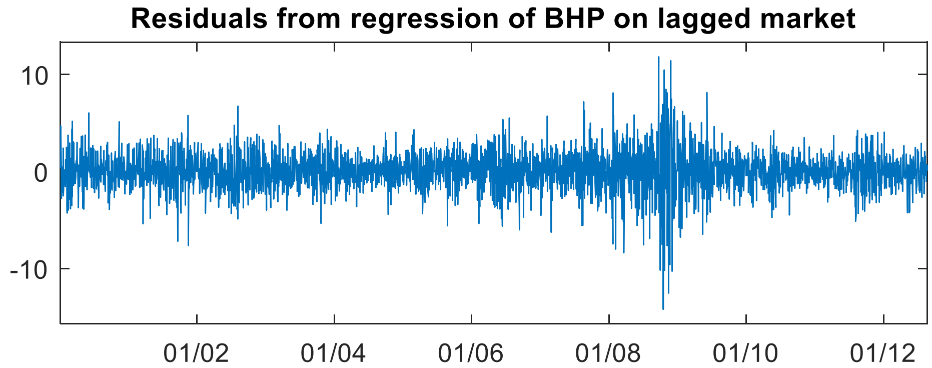


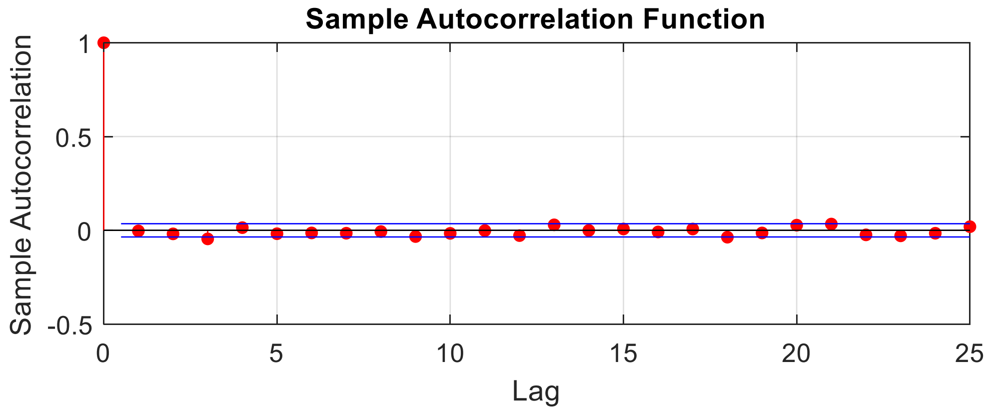
Where the 3rd correlation seems marginally significant, hence I chose an AR(3), estimated as:

$BHP_t = 0.06 - 0.0408\ BHP_{t-1} - 0.0202\ BHP_{t-2} - 0.0504\ BHP_{t-3} + \varepsilon_t$
SER = 2.01%

p-values on LB tests for 5 and 10 lags of autocorrelation equalling 0 were 0.0091 and 0.0211 respectively, thus significant autocorrelation does exist in the first 5, but possibly not in first 10 lags of BHP returns.

The residuals from the regression model (no. 4) for BHP are plot over time and with ACF





p-values on LB tests for 5 and 10 lags of autocorrelation (with 4 and 9 degrees of freedom) equalling 0 were 0.05 and 0.08 respectively, thus significant autocorrelation may exist in the first 5 lags of BHP residuals, but perhaps not in the first 10 lags as a group. Observing the marginally significant 3rd lag in the ACF plot, I chose a Reg-AR(3), estimated as:

$BHP_t = 0.060 - 0.085 AORD_{t-1} - 0.009\ BHP_{t-1} - 0.019\ BHP_{t-2} - 0.048\ BHP_{t-3} + \varepsilon_t$

SER = 2.00%

For the last model I added a 5th lag, being for the same day the week before.

$$BHP_t = 0.061 - 0.082 AORD_{t-1} - 0.009\ BHP_{t-1} - 0.019 BHP_{t-2} - 0.048 BHP_{t-3} + 0.012\ BHP_{t-4} - 0.018 BHP_{t-5} + \varepsilon_t$$

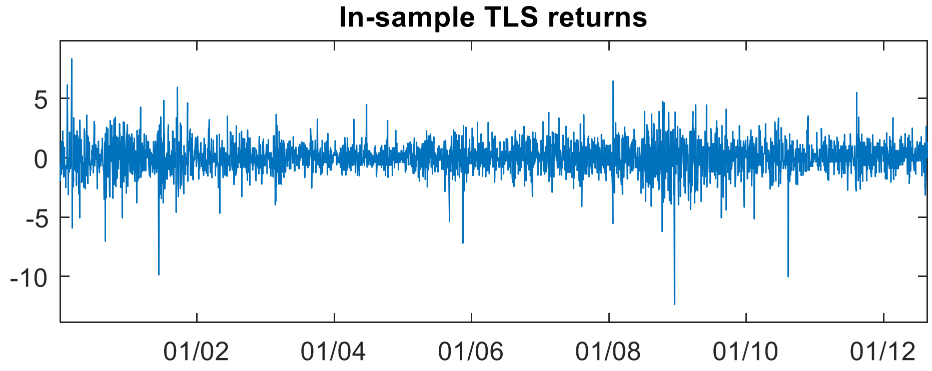
SER = 2.00%

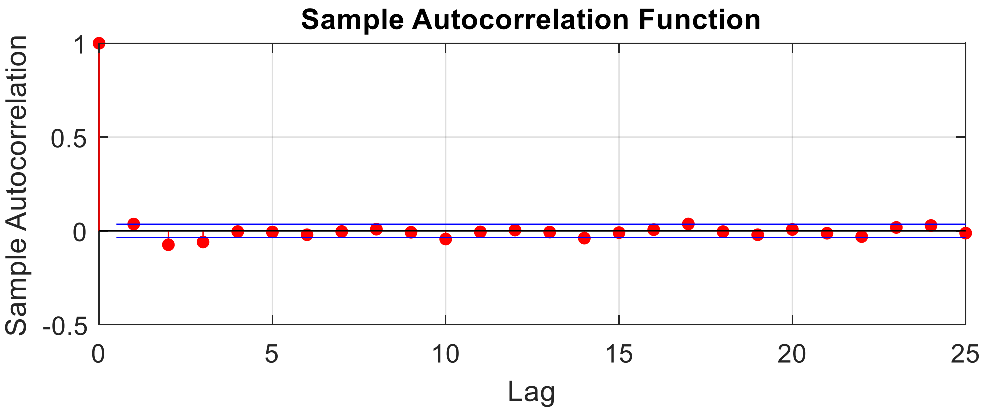
<u>For TLS</u>

The regression model is estimated as:

$TLS_t = 0.0144 - 0.0624 AORD_{t-1} + \varepsilon_t$; $R^2 = 0.22\%$, $SER = 1.35\%$

An ACF plot for TLS returns shows



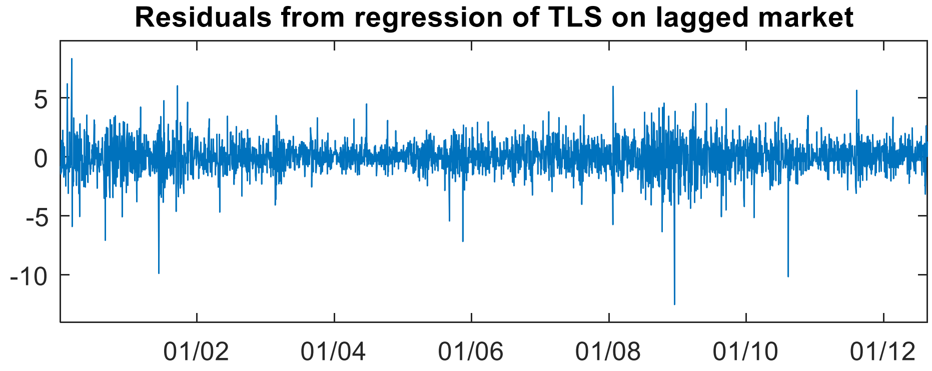


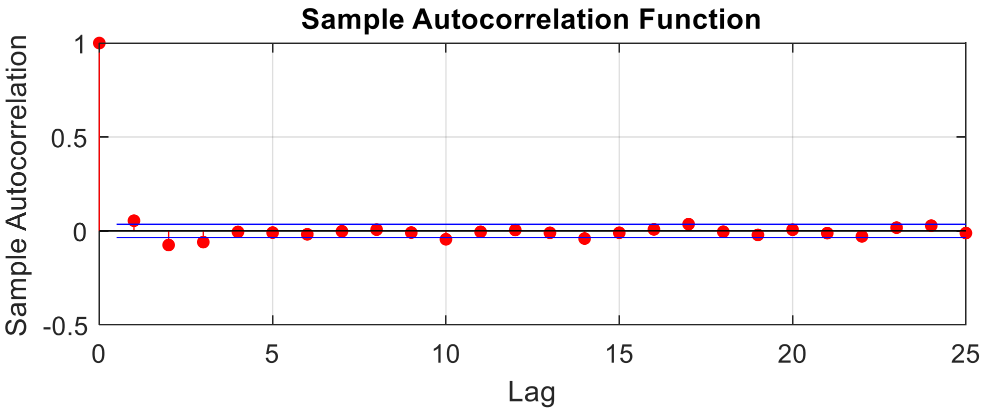
The 2nd and 3rd correlations seem quite significant. Furthermore, the p-values on LB tests for 5 and 10 lags of autocorrelation equalling 0 were 0.32*e-05 and 0.97*e-05 respectively, thus significant autocorrelation does exists in the first 5 and 10 lags of TLS returns.
Hence I chose an AR(3), estimated as:

$TLS_t = 0.014 + 0.035\, TLS_{t-1} - 0.073\, TLS_{t-2} - 0.054\, TLS_{t-3} + \varepsilon_t$

SER = 1.35%

The residuals from the regression model (no. 4) for TLS are plot over time and with ACF





p-values on LB tests for 5 and 10 lags of autocorrelation (with 4 and 9 degrees of freedom) equalling 0 were almost 0, thus significant autocorrelation does exist in the first 5, and 10 lags of TLS residuals. Observing the significant 2nd and 3rd lags in the ACF plot, I chose a Reg-AR(3), estimated as:

$$TLS_t = 0.0163 - 0.099AORD_{t-1} + 0.064TLS_{t-1} - 0.076TLS_{t-2} - 0.053TLS_{t-3} + \varepsilon_t$$

SER = 1.35%

For the last model I added a 5th lag, being for the same day the week before.

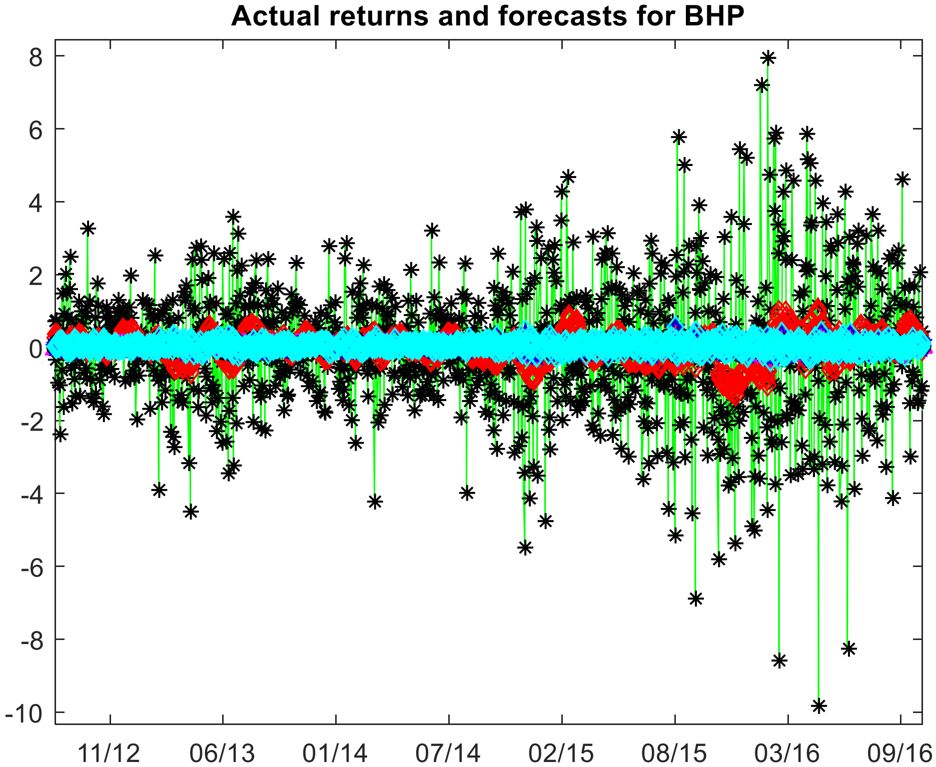
$$TLS_t = 0.017 - 0.099AORD_{t-1} - 0.064TLS_{t-1} - 0.077TLS_{t-2} - 0.054TLS_{t-3}$$
$$- 0.006TLS_{t-4} - 0.016TLS_{t-5} + \varepsilon_t$$

SER = 1.35%

(c) *Generate moving origin horizon 1 forecasts for each observation in your forecast sample from all methods above. Assess the accuracy of these forecasting methods using plots, RMSE and MAD.*

The symbols used for each method are the same in all the plots below. None of the forecasts seem to “follow” the directions or magnitudes of the actual returns they are trying to forecast. This pattern is repeated for both assets.

For BHP

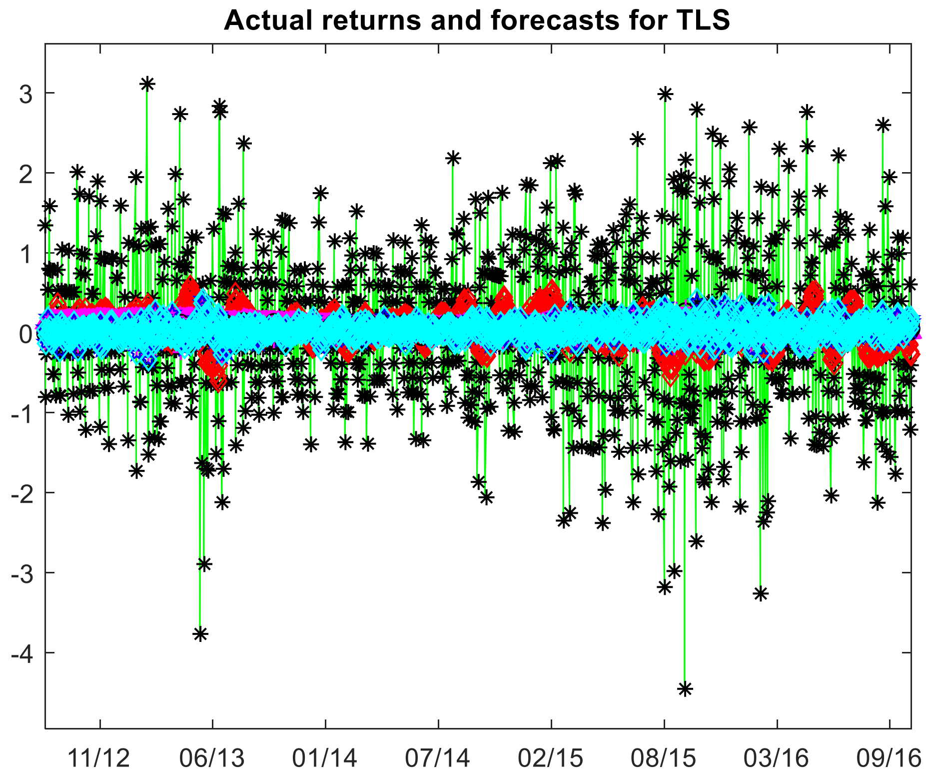


Forecast accuracy

	Naive	1 mth	12 mth	Reg	AR(3)	Reg-AR(3)	Reg-AR(5)
	1	2	3	4	5	6	7
RMSE	*2.4618*	1.8411	**1.8021**	1.8023	1.8046	1.8031	1.8041
MAD	*1.8177*	1.3419	**1.3159**	1.3180	1.3209	1.3195	1.3203

The units for these measures are the same units as percentage returns. The typical errors made are between 1.3% and 2.5% in terms of percentage daily returns. These seem somewhat large, especially compared to the standard deviation, which is 1.99! In fact almost all models do better than those measures, indicating that all these models work to a degree. The best method under RMSE the 12 mth mean closely followed by the Reg and AR methods. Under MAD the 12 month model is again the most accurate. Again there is little difference between the AR models and the Reg model. The naive method ranks last under both measures and the 1 month mean is second last under both measures. Under both RMSE and MAD the regression method is second best.

For TLS



Forecast accuracy

	Naive	1 mth	12 mth	Reg	AR(3)	Reg-AR(3)	Reg-AR(5)
	1	2	3	4	5	6	7
RMSE	*1.3664*	0.9712	0.9492	**0.9488**	0.9497	0.9504	0.9512
MAD	*1.0360*	0.7510	0.7346	0.7338	**0.7330**	0.7341	0.7354

The units for these measures are again the same units as percentage returns. The typical errors made are between 0.73% and 1.37% in terms of percentage daily returns. These again seem somewhat large, especially compared to the standard deviation, which is 1.1988. It seems like all models except from Naive model work to a degree. The best method under RMSE is the Reg method and following MAD the AR(3) is most accurate. Again the naive method ranks last under both measures.

(d) Comment on the results obtained regarding forecast accuracy over all methods.

For BHP the 12-month mean performed best whereas for TLS the regression methods performed slightly better. In terms of the regression methods it seems like the lagged market index is the best predictor overall for both stocks. However, the differences in scores are very small, indicating no forecasting model is significantly better than others.

The worst method regarding accuracy is always the naive. This would do well if the returns were behaving like a random walk over time. Since they instead are mean stationary, with a mean close to 0, this method does very poorly in accuracy. (Remember that market equilibrium arguments would suggest that prices follow a random walk, but not the returns).