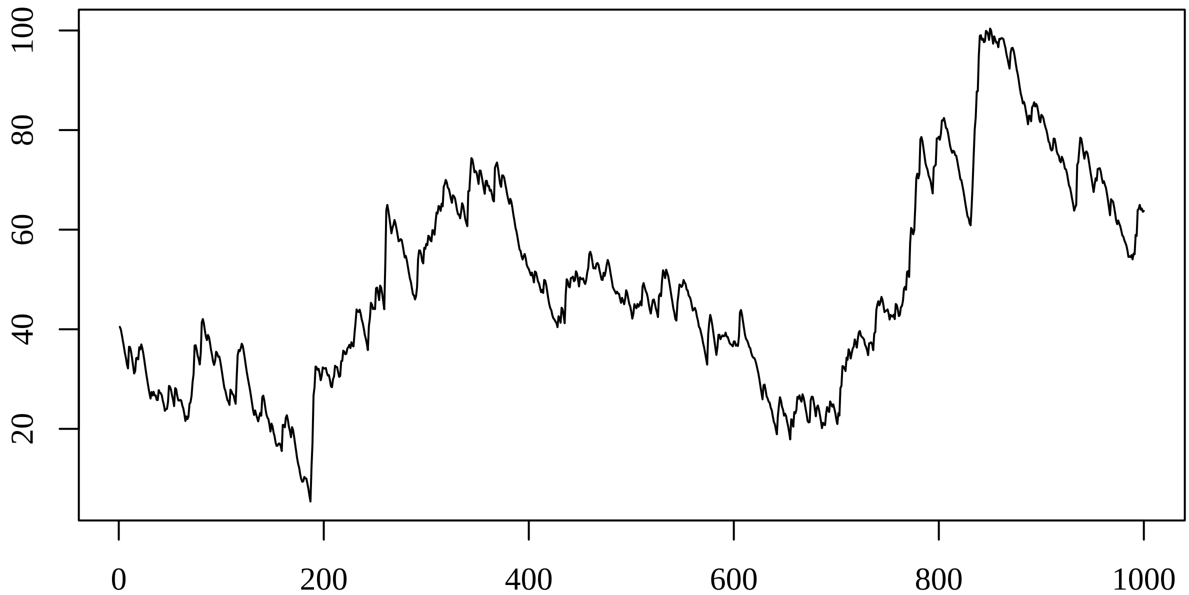
Question 5.

The following time series plot show $n = 1,000$ observations on an asset price P_t. The returns calculated from this price series are calculated ($Y_t = \Delta \log P_t$) and an AR(1)-GARCH(1,1) model is fitted.



The coefficient estimates are as follows:

```
   mu    ar1  omega alpha1  beta1
0.025  0.270  0.785  0.088  0.593
```

To be specific, `mu` is the estimate of the mean of Y_t, `ar1` the coefficient of the AR(1) equation, and `omega`, `alpha1` and `beta1` the coefficients of the GARCH(1,1) equation.

(a) Write out the estimated model in equation form.

The estimated model can be written

$$
\begin{aligned}
Y_t &= 0.025 + \widehat{Z}_t \\
Z_t &= 0.270\, Z_{t-1} + \widehat{U}_t \\
\widehat{\sigma}_t^2 &= 0.785 + 0.088\, U_{t-1}^2 + 0.593\, \sigma_{t-1}^2
\end{aligned}
$$

where $\sigma_t^2 = \mathrm{var}(U_t | \mathscr{Y}_{t-1})$.

The following statistics were computed from the standardised residuals of the model.

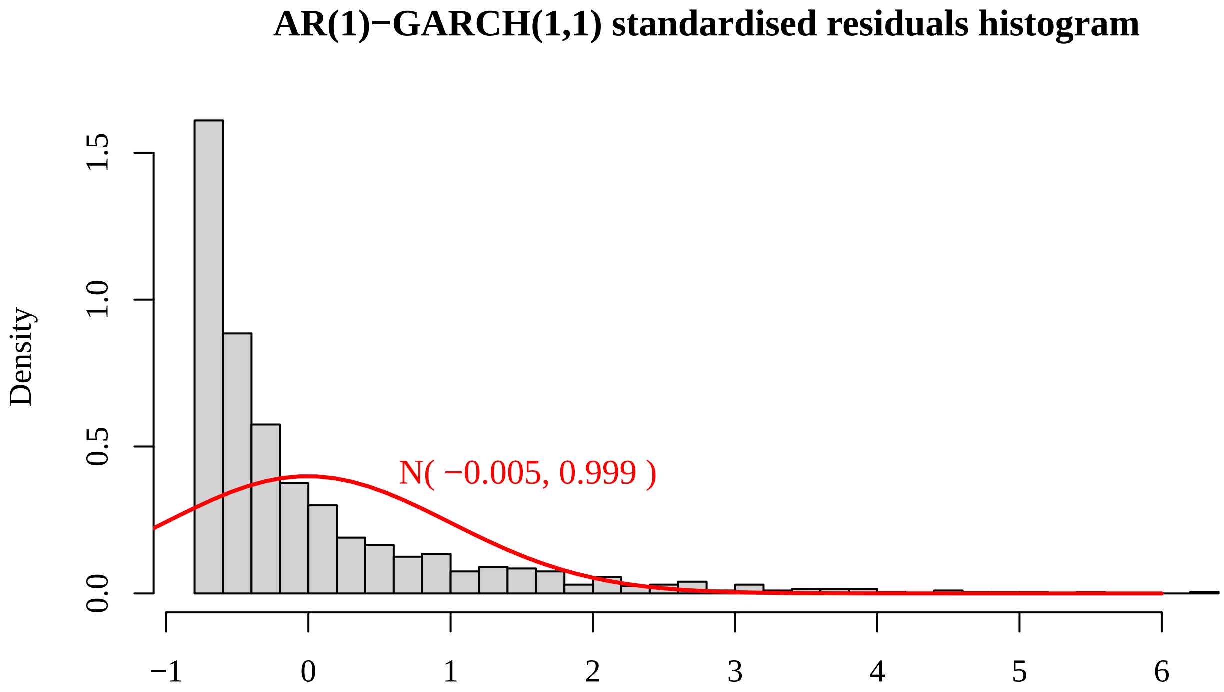
```
Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic p-value
Lag[1]                    0.08651  0.7687
Lag[2]                    0.97189  0.7567
Lag[5]                    1.76399  0.7779
H0 : No serial correlation
Weighted ARCH LM Tests
------------------------------------
            Statistic Shape Scale P-Value
ARCH Lag[3] 0.0004424 0.500 2.000  0.9832
ARCH Lag[5] 0.4603905 1.440 1.667  0.8953
ARCH Lag[7] 0.5228336 2.315 1.543  0.9763
```

(b) What do these tests suggest about the specification of the model? Explain.

The Ljung-Box tests are for autocorrelation in the residuals. Autocorrelation in the residuals would imply the model for the mean (the AR(1) model) is misspecified. In this case the tests at lags 1, 2 and 5 have p-values above 0.05, implying the the null hypothesis of no autocorrelation is not rejected in each case. There is no evidence here of misspecification in the mean equation.

The ARCH LM tests are for conditional heteroskedasticity remaining in the residuals, which would imply the GARCH(1,1) model is misspecified. In this case the tests at lags 3, 5 and 7 have p-values above 0.05, implying the null hypothesis of conditional homoskedasticity is not rejected. There is no evidence of misspecification in the variance equation.

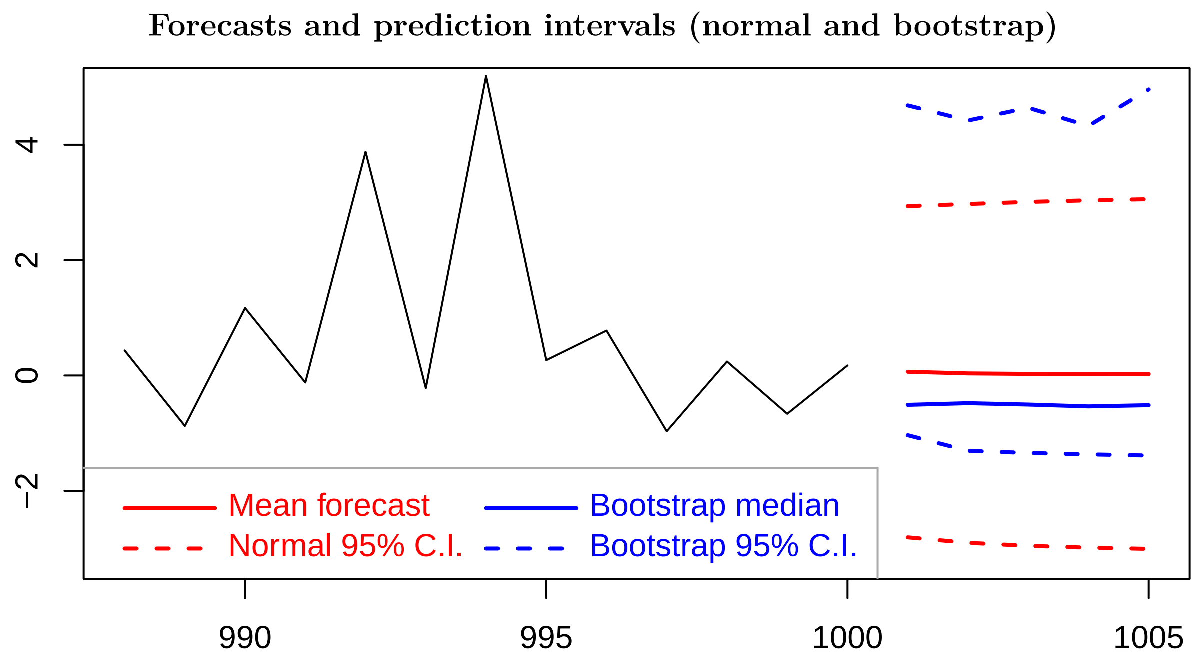
Following are some graphical and descriptive statistics relevant to the distribution of the standardised residuals.



Mean	Std.Dev.	Skewness	Kurtosis
−0.005	1.000	2.318	9.386

(c) What do you conclude about the normality or otherwise of the standardised residuals?

There is clear evidence of substantial deviations from normality. The standardised residuals have considerable positive skewness excess kurtosis. The histogram illustrates both of these characteristics.



The plot above and output table below show up to 5-step-ahead forecasts and prediction intervals from the model. The usual conditional mean point forecasts are denoted `Mean`, with prediction intervals computed assuming conditional normality denoted `L_Norm` and `U_Norm`. The bootstrap was also applied to produce 5,000 replications over the forecast period, with the 2.5%, 50% and 97.5% quantiles of these replications denoted `L_Boot`, `Median` and `U_Boot` respectively.

h	L_Norm	L_Boot	Median	Mean	U_Norm	U_Boot	Sigma
1	-2.806	-1.035	-0.509	0.065	2.936	4.682	1.465
2	-2.901	-1.306	-0.480	0.036	2.973	4.420	1.499
3	-2.953	-1.343	-0.504	0.028	3.009	4.641	1.521
4	-2.985	-1.367	-0.536	0.026	3.037	4.327	1.536
5	-3.006	-1.388	-0.515	0.025	3.056	4.959	1.546

(d) What are the interpretations of the two confidence intervals? Why do they differ, and which would you prefer for this application?

Both intervals are constructed to attempt to include the actual value with probability 95% at each forecast horizon.

The normal intervals are constructed as `Mean`$\pm$1.96 `Sigma`, a formula which assume conditional normality of the dependent variable.

The bootstrap intervals are constructed by repeatedly sampling from the residuals of the model and using these to construct simulated paths for the forecast period. The quantiles of these bootstrap samples are then used for the intervals. There is no assumption of normality, the bootstrap is instead drawing from the observed distribution of the residuals.

Since the residuals are very clearly far from normal, the bootstrap intervals may be considered preferable in this case. (It may also be noticed in this case that the bootstrap intervals correspond more closely with the range of observations in the sample, while the normal intervals appear to be shifted downwards relative to the observations.)

(e) Return to the model in equation form in part (a). Show how this can be rearranged to produce estimated equations for $\widehat{E}(Y_t|\mathscr{Y}_{t-1})$ and $\widehat{\text{var}}(Y_t|\mathscr{Y}_{t-1})$ written in terms of lags of the observed time series Y_t on the right hand side (i.e. not lags of Z_t or U_t) which could be used for computational purposes.

$$\widehat{E}(Y_t|\mathscr{Y}_{t-1}) = 0.025 + 0.270\,(Y_{t-1} - 0.025)$$

$$\widehat{\text{var}}(Y_t|\mathscr{Y}_{t-1}) = 0.785 + 0.088[(Y_{t-1} - 0.025) - 0.270(Y_{t-2} - 0.025)]^2 + 0.593\,\widehat{\text{var}}(Y_{t-1}|\mathscr{Y}_{t-2})$$