Report - SF2943 Time Series Analysis

Group TS 6: Buqing Cao Titing Cui Xin Huang Zeyu Chen

May 16, 2018

Part 1 Peer Review (SF2943 BKMP)

First, in this well-organised report, the Augmented Dickey-Fuller test is applied to determine the differencing order, which follows the recommended work flow introduced in the course. Also, the ACF values are presented for preliminarily differenced series, which provided a validated evidence for suitability of model assumptions. Further, the well-acknowledged AIC and BIC are implemented in the model selection process, which gives a quite convincing result of ARMA(1,2) model.

There are just a few small problems. From the graph of the original data, we guess that a quadratic trend may seem to be more suitable. Especially after 1980, a growing trend is more significant. Therefore, considering a differencing with order = 2, i.e. $(1-B)^2$ may be helpful. Another method that may be applicable is to divide the data set into two parts. Besides, adding some tests of the residuals may further increase credibility . The diagnostics of the residuals can be performed with more applicable methods. A qq-norm test or Ljung-Box test may provide more supportive information for their model.

Part 2 Literature Review (Article 10)

(1) Recapitulation of findings of the research article

Based upon a data set consisting of hourly average wind speed measurements collected on five weather stations in Navarre (Spain) for nine years, the authors conducted a thorough and comprehensive analysis on these time series with ARMA model, and showed a considerable improvement by comparison with a relatively simple Persistence model.

The studies can be summarized with the three following procedures.

First, according to a preceding research, hourly wind speed can be better adjusted to a Weibull distribution, which may cause the non-Gaussian property, therefore a transformation with exponent m is implemented by an iterative calculation process. Then a standardization is carried out, with the formula:

$$V_{n,y}^{*} = \frac{V_{n,y}^{'} - \mu(t)}{\sigma(t)}$$

where $\mu(t)$ and $\sigma(t)$ are mean value and standard deviation of the transformed data for each hour, $1 \leq t \leq 24$. In order to avoid seasonality problems, the authors adjusted a different model for each calendar month.

Second, in order to apply the ARMA(p,q) model on these time series, a process of three phases is performed.

(i) Identification Phase

In order to determine the indices p and q of the ARMA model, the authors firstly computed the sample ACF and PACF values. Then combining the information obtained from the following parameter estimation phase, applying the identification method proposed by Tsay and Tiao (1984) and Beguin et al. (1980), an adequate group of models are contrasted with the BIC or AIC criterion.

(ii) Parameter Estimation Phase

In this phase, firstly a preliminary estimation is done by applying the Yule-Walker relations for the auto-regression coefficients, and the Newton-Raphson algorithm (propose by Box and Jenkins, 1976) for the moving average coefficients. Then using the algorithm proposed by Marquardt, by doing a forecast backwards and minimizing the sum of squares of residuals, the final coefficients are determined.

(iii) Validation Phase

A global contrast is applied using the statistic proposed by Box-Pierce (1970), if it is not supported that this statistic follows a χ^2 distribution, we have no sufficient reason to reject the Null hypothesis and the model may not be accepted. Finally 90% of the total sixty series satisfied the Box-Pierce criterion at a level of 0.1.

Lastly, the authors made a forecast from the discovered ARMA(p,q) models and assessed their degree of success, by computing the root mean square error (RMSE) and mean absolute error (MABE) from comparison between the predicted values and the actual values. For relatively short-term forecasts, the RMSE can be within the 1.5 m/s threshold.

(2) Cogency of the analysis

(i) Relevance

By computing the RMSE (averaged over all months) for the prediction from ARMA and Persistence model, the authors successfully showed that ARMA models can provide considerable improvement in the accuracy of forecast.

(ii) Originality

From a more practical point of view, the authors made an analysis of the RMSE values of forecast, in different wind velocities. Thus it will be convenient to check the forecast accuracy for the range of wind velocities in which the energy generators may be more interested.

(iii) Accuracy of analysis

From the Autocorrelograms of the residuals obtained by applying the ARMA models selected, we can observe that the ACF values of residuals are all very small in value, which can increase our confidence in the accuracy of these models.

(iv) Suitability of model assumptions
Of the total 60 time series, 54 of their corresponding ARMA models can pass the Box-Pierce test, suggesting that ARMA models could be suitable in a majority of situations.

(3) Possible aspects for improvement

(i). In order to address the non-stationarity property of the original time series, it may also be an applicable idea to implement differencing method with suitable orders.

(ii). According to our course textbook, a distinctive feature of slowly decaying positive sample autocorrelation function suggests the appropriateness of an ARIMA model. In fact this is just the case shown in the sample ACF plot of this article. Perhaps we can also try to fit the time series with an ARIMA model.

Part 3 Analysis of financial time series

a1)

(i) With 'today' (t=0) setting as Jan 11th 2018, firstly we can follow the naive approach to calculate the corresponding 0.05-quantile for the value S_1. With equation (1):

$$S_k - S_0 = S_0(e^{X_1+\ldots+X_k} - 1) \overset{\mathrm{d}}{\approx} S_0(e^{k\hat{\mu}+\sqrt{k}\hat{\sigma}Z-1})$$

where Z is a standard normal variable. As we are suggested to use the last n observations for volatility estimator, we used the data from 2017-08-31 to 2018-01-11, satisfying the requirement of 50 $<n<$100. From the workspace, we can obtain the estimation for parameters as: $\hat{\mu} = 0.00060379$, $\hat{\sigma} = 0.00588005$, and also that $S_0 = 1630.29$. Since we know that the 0.05-quantile of standard normal variable $Z_{0.05} = \Phi^{-1}(0.975) \approx 1.96$, the corresponding 0.05-quantile of S1 can be computed as:

$$S_0(e^{\hat{\mu}+\hat{\sigma}Z_{0.05}}) \approx 1650.18$$

Secondly, we can try with a GARCH(1,1) model to fit the same data set, where we can simply assume that $\mu = 0$. Since k = 1, we can have:

$$\sigma_1^2 = \alpha_0 + \alpha_1 X_0^2 + \beta_1 \sigma_0^2$$
$$X_1 = \sigma_1 Z_1$$
$$S_0 = S_1 e^{X_1}$$

From the workspace we can know the estimation of σ_1 is about 0.005070, thus we can compute the 0.05-quantile of S_1 as:

$$S_0(e^{\hat{\sigma_1}Z_{0.05}}) \approx 1630.29 \times exp(0.005070 \times 1.96) \approx 1646.57$$

(ii) Similarly with (i), the dates for data set are chosen as: from 2017-09-30 to 2018-02-09. The present value $S_0 = 1500.18$, and for the naive model, we have parameters as: $\hat{\mu} = -0.00104076$, $\hat{\sigma} = 0.00788848$. The corresponding 0.05-quantile of S_1 can be computed as:

$$S_0(e^{\hat{\mu}+\hat{\sigma}Z_{0.05}}) \approx 1521.97$$

Also applying a GARCH(1,1) model on the same data set, we can obtain the estimated σ_1 value as 0.005273, and the 0.05-quantile of S_1 can be given from:

$$S_0(e^{\sigma_1 Z_{0.05}}) = 1500.18 \times exp(0.005273 \times 1.96) \approx 1515.77$$

a2)

(i) For today setting as Jan 11th, 2018, also using the same historical data set as in a1), we can still obtain the same estimated parameters for the naive mode. In order to look five days ahead, we can apply equation (1) and obtain the 0.05-quantile of S_5 as:

$$S_0(e^{k\hat{\mu}+\sqrt{k}\hat{\sigma}Z_{0.05}}) \approx 1630.29 \times exp(5 \times 0.00060379 + \sqrt{5} \times 0.00588005 \times 1.96) \approx 1677.91$$

In order to implement a simulation with GARCH(1,1) model, we used the coefficients from the workspace as: $\alpha_0 = 0.000017288$, $\alpha_1 = 0.098709121$, $\beta_1 = 0.363079276$, $\sigma_1 = 0.00507$ and $X_1 = \sigma_1 Z_1$, Z_1 is a standard normal random variable. Then from the scheme of GARCH(1,1) model,

$$X_t = \sigma_t Z_t$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$
$$S_5 = S_0 e^{(X_1+\ldots+X_5)}$$

We can obtain a simulation for S_5 value by iterating this procedure for 4 times. Further by independently constructing this process by N = 10^4 times, we can finally receive a sample of size N for S_5 values, and the empirical 0.05-quantile for these samples are given as:

$$S_5(0.05) \approx 1670.19$$

(ii) For today setting as Feb 9th, 2018, with the same method as in (i), firstly applying the naive model, we can obtain $S_5 = S_0(e^{k\hat{\mu}+\sqrt{k}\hat{\sigma_1}Z_{0.05}}) \approx 1544.89$, where the parameters are still given as: $\hat{\mu} = -0.00104076$, $\hat{\sigma} = 0.00788848$, and also that $S_0 = 1500.18$.

Also following the same simulation algorithm, we can obtain the empirical 0.05-quantile of 10000 sample S5 values can be given as $S_5(0.05) \approx 1546.03$.

b)

Take the situation of 2018-01-11 for an example, we can also compute the lower-bound 0.05-quantile and the corresponding 95% confidence intervals of S_1 for two different models as:

Model	Upper 0.05-quantile	Lower 0.05-quantile	95% CI
Naive Model	1650.18	1612.59	(1612.59, 1650.18)
GARCH Model	1646.57	1614.17	(1614.17, 1646.57)

The length of the confidence interval for the naive and GARCH models are respectively 37.59 and 32.40, suggesting that GARCH model can provide a relatively more accurate forecast. Also from the graph of the log-returns together with the GARCH-implied and the naive volatility, we can also observe that when there are large fluctuations in the log-return data, GARCH model tend to be followed by fluctuations of comparable magnitude. We can see that the GARCH model is suitable to characterize the features of the true log-returns.

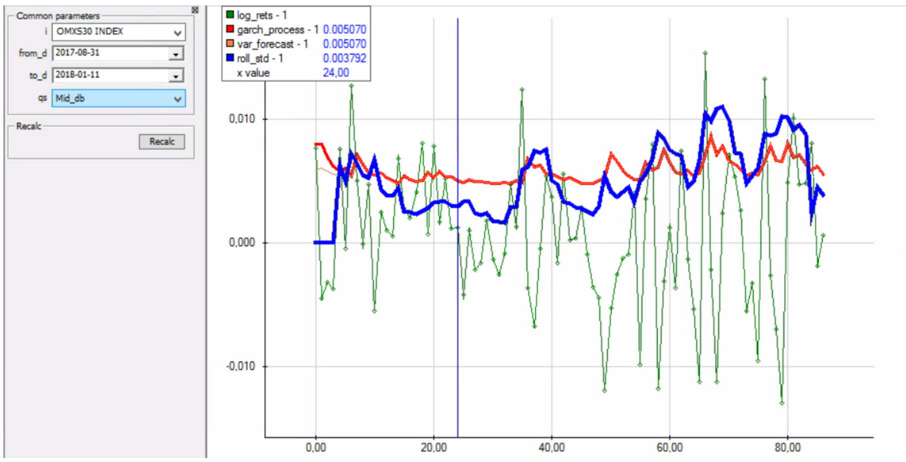

Figure 1: The log-returns together with the GARCH-implied and the naive volatility

The observation of two models for 2018-02-09 also shows similar behavior, the length of confidence intervals for naive and GARCH model are 46.34 and 31.01 respectively, still verifying an improved accuracy of GARCH model compared to the naive model.

References

[1] Brockwell, P. J. and Davis R. A. *Introduction to Time Series and Forecasting*, 2nd edition, 8th corrected printing. Springer, 2010.