MSc Mathematical and Computational Finance

Statistics and Financial Data Analysis - Problem Sheet 5

Please restrict answers to Questions 1 and 2 to 2 pages each, remembering to write your main observations and conclusions, using data analysis, statistical tests (with their interpretation) and appropriately labelled plots to support your findings.

1. Perform a survey of financial assets, and their autoregressive structure, on a daily timescale, over a period of two years (2018-2019). You should discuss:

    - whether the log returns series is normally distributed
    - whether there is a significant autocorrelative structure in the returns and squared returns at one or two lags.

    Your survey should include most constituents of the SP500 saved in a CSV file: '`SP500.csv`'. Present your results in an informative way (a table of all 500 constituents of the SP500 is not desired). Do you have more significant autocorrelations than would be the case under a null hypothesis?

    A modification of the following script, which reads in the `Symbol` column from the file '`SP500.csv`' may be convenient (note that running this script may take a few minutes). Once the data is downloaded, you can call the `acf` function from `statsmodels.tsa.stattools` and store results for lag 1 and 2 of each component.

```
start_date = '2018-01-01'
end_date = '2019-12-31'

allData = pd.DataFrame()

for i in range(nSpy):
    tick = df['Symbol'][i]
    print(tick)
    try:
        temp = pd_data.DataReader(tick,'yahoo',start_date,end_date)['Adj Close']
        if(len(temp)<500):
            continue
        else:
            allData[tick] = temp
    except:
        print('Failed: ', tick)

        continue
```

2. Consider the growth in manufacturing productivity in the UK, an index of which has been downloaded from the following site and saved under the CSV file name '`mfp.csv`':

    http://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/labourproductivity/timeseries/a4ym/prdy

    Using the quarterly data, investigate the possible autoregressive structure of the **growth** in productivity. Using the data provided, what predictions for the increase in productivity it estimate for 2019-20? (use data up to Q2 2018 as in the CSV file). How has productivity been changing over the long-term?

By re-fitting your model over the periods 1997–2007 and 2008–2018, discuss what apparent changes have been seen in productivity growth following the 2008 financial crisis. Are these changes significant (both economically and statistically)? How does this change your predictions of productivity growth?

N.B. The following version of the t-test may prove useful. Suppose for populations $i = 1, 2$ you have independent estimates $\hat{x}_i$ of a parameter μ_i (e.g. the mean in an ARMA model) along with a squared standard error S_i^2 (i.e. standard errors of the estimator $\hat{x}_i$, not the standard deviation of the population), sampled with degrees of freedom d_i respectively. Recall that the appropriate degrees of freedom for an estimate in a regression is $n - k$, where k is the number of parameters fitted before calculating the variance. For an ARMA model, the estimated coefficients can be found with fit.params and their covariance matrix with fit.cov_params(), which has the squared standard errors on the diagonal (or by looking at the summary when fitted using ARMA in the statsmodels package). Then to test whether the parameter is the same in each model (without assuming they have the same standard errors) we can use the test statistic:

$$t = \frac{\hat{x}_1 - \hat{x}_2}{\sqrt{s_1^2 + s_2^2}}$$

Under the assumption that $\mu_1 = \mu_2$, t has an approximate t-distribution, with degrees of freedom ν given by the Welch–Satterthwaite approximation:

$$\nu = \frac{(s_1^2 + s_2^2)^2}{s_1^4/d_1 + s_2^4/d_2}$$

3. Consider an unnormalised AR(1) process, that is, where $X_t = \phi X_{t-1} + Z_t$, ϕ is a constant and $Z \sim WN(0, \sigma^2)$.

   (a) For $|\phi| < 1$, given that $X_0 \sim N(\mu, \kappa^2)$, give a formula for the distribution of X_t. Hence show that there exist values μ, κ such that X is a weakly stationary process, when Z_t is independent of X_{t-1} for each t.

   (b) Show that, if $|\phi| < 1$, then X_t converges in distribution to the stationary distribution, as $t \to \infty$. (You may use the fact that, for a sequence of normal distributions, the means and variances converge if and only if the distributions converge.)

   (c) By reversing time, show that, if $|\phi| > 1$, then there is also a stationary solution (i.e. you can choose μ, κ and the correlation between X and Z to make X stationary). What is the covariance between Z_t and X_{t-1} in the stationary solution?

   (d) Give an argument why, if $|\phi| > 1$, then the stationary solution is unstable (i.e. if you start anywhere except at the stationary solution, the distribution will not converge).