School of Electronics and Computer Science
University of Southampton

COMP6212 Computational Finance 2017/18

Assignment (Part I (2 & 3 (merged) of 3 from MN): 30%)

Issue	11 March 2019
Due	26 March 2019 (15:00)

Please work on Section 1 [20 marks] and either one of sections 2 or 3 [10 marks each].

1 Derivatives Pricing [20 Marks]

Some prices of call and put options written on the `FTSE100` Index are available in the course notes page (file `FTSEOPtionsData`). These are daily data during the period January 2018 to January 2019, all of which are European style exercise and maturing in January 2019. Select parts of this data for the exercises below.

a Study how the Black-Scholes model of pricing options was derived before attempting this task. The notation is : K, the strike price; S, the value of the underlying asset; r the risk-free interest rate; T, the time of maturity; t, the current time and $\mathcal{N}(x)$, the cummulative normal distribution.

1. Write the expression for $\mathcal{N}'(x)$ (the derivative of $\mathcal{N}(x)$).
2. Show that $S\,\mathcal{N}'(d_1) = K\,\exp(-r(T-t))\,\mathcal{N}'(d_2)$ where d_1 and d_2 were defined as:

$$\begin{aligned} d_1 &= \frac{\log(S/K) \;+\; (r+\sigma^2/2)(T-t)}{\sigma\sqrt{(T-t)}} \\ d_2 &= d_1 \;-\; \sigma\sqrt{(T-t)} \end{aligned}$$

3. Calculate the derivaties $\partial d_1/\partial S$ and $\partial d_2/\partial S$
4. With the solution for the call option price given by

$$c \;=\; S\mathcal{N}(d_1) \;-\; K\exp(-r(T-t))\,\mathcal{N}(d_2)$$

show that its derivative with respect to time is

$$\frac{\partial c}{\partial t} \;=\; -rK\exp(-r(T-t))\,\mathcal{N}(d_2) - S\mathcal{N}'(d_1)\frac{\sigma}{2\sqrt{(T-t)}}$$

Show that $\partial c/\partial S \;=\; \mathcal{N}(d_1)$

5. Differentiating again to get

$$\frac{\partial^2 c}{\partial S^2} \;=\; \mathcal{N}'(d_1)\frac{1}{S\sigma\sqrt{(T-t)}},$$

and substituting in the relevant expression, show that the expression for the call option price indeed the solution to the Black-Scholes differential equation.

b Evaluate how well the option prices in the given data satisfy the Black-Scholes model. For each option, evaluate the price given by Black-Scholes from $T/4+1$ to T (the length of the time series), using volatility estimated using a sliding window of range $t - T/4$ to t. Compare the price obtained by using the formula with the true price of the option. Are there any systematic differences? For estimating volatility from historic data, see for example Hull [1] Section 13.4 or a similar source.

c On a random set of 30 days in the range $T/4 + 1 : T$, compute the *implied volatilities* and plot them as scatter plot against the corresponding volatilities you estimated from data. For prices on any particular day, is there any systematic variation of implied volatilities computed from options with different strike prices (Hint: Look up the term *volatility smile*)?

d Compare pricing a call option with European style exercise using the Black-Scholes and Binomial lattice methods. Using *one* random set of values for the parameters (strike price, time to maturity, interest rate and volatility) taken from the data given to you, evaluate how the binomial lattice method approximates Black-Scholes as the step time δt is decreased. Plot a graph of the absolute difference between the two methods as a function of δt.

Consider pricing an option using a binomial lattice. The code for pricing a put option with American style exercise included the following lines[2]:

```
[...]
for tau=1:N
    for i= (tau+1):2:(2*N+1-tau)
        hold = p_u*PVals(i+1) + p_d*PVals(i-1);
        PVals(i) = max(hold, K-SVals(i));
    end
end
[...]
```

Explain *in your own words* the steps involved in this part of the code. How will this change if you were pricing a call option (with American style exercise)?

2 Data-driven Modelling of Option Prices [10 Marks]

The relationship between the price of an option contract and the underlying parameters relating to the exercise of the option is a complex one. Hutchinson *et al.* [3] show that this relationship can be reasonably approximated by nonparametric neural network type models. In particular, they train a Radial Basis Functions (RBF) model on data simulated from a Black-Scholes model and show that not only good approximations to the options price can be obtained, but also the sensitivity *Delta* ($\Delta = \partial C / \partial S$) can be relaibly extracted from it. In this assignment we will explore if some of the claims made in their paper are true.

1. Study the paper by Hutchinson *et al.* [3]. You need to focus on Figures 4 and 5, and how these were arrived at. Skip their evaluation part beyond this point.

2. Construct a dataset by generating call option prices from the Black-Scholes formula. You may take input parameters (strike prices, interest rates, volatilities and the underlying asset price) from the data used in Assignment 2.

   With this data, reproduce Fig. 4 and Fig. 5 in [3] by training an RBF model (Eqn. 9 in [3]).

   - With the data vector $\boldsymbol{x} = [\, S/X \quad (T-t)\,]^T$, the RBF model used is

   $$c = \sum_{j=1}^{J} \lambda_j \, \phi_j(\boldsymbol{x}) \, + \, \boldsymbol{w}^T \boldsymbol{x} \, + \, w_0$$

     where the nonlinear function $\phi_j(\boldsymbol{x}) = \left[(\boldsymbol{x} - \boldsymbol{m}_j)^T \, \boldsymbol{\Sigma_j} \, (\boldsymbol{x} - \boldsymbol{m}_j) \, + \, b_j \right]^{1/2}$ is a local Mahalanobis distance with a bias term and $J = 4$. In this task, we will ignore the bias terms within the nonlinearity (*i.e.* $b_j = 0$).
   - Use the MATLAB function `fitgmdist` for fitting a Gaussian mixture model and extract the parameters $\boldsymbol{m}_j$ and $\boldsymbol{\Sigma}_j$ of the nonlinear terms in the model.
   - Now construct the *design matrix* of inputs – this matrix will have as many rows as you have data points and seven columns (four for mapping the data to each of the nonlinear distances, two for the two data dimensions of the linear part and the last column will have ones which multiply w_0).

- Estimate the weights of the model ($\lambda_j, j = 1, ..., 4$; $\boldsymbol{w}$ and w_0) by solving the resulting least squares problem.

3. How well does such a non-parametric model learn the data used in Section 1?

4. Compute the "hedge ratio" (*i.e.* the derivative of option price with respoect to the underlying asset price) by differentiating the network and from Black-Scholes formula. How well do they compare?

3 Time Series Analysis [10 Marks]

In a recent paper, Mahler [4] shows how the Kalman filter for sequential estimation of a state space model can be combined with the *Lasso* approach [5] for l_1 penalized regression. The result is to select lags (past values) of explanatory variables that are relevant for the prediction of future values of a time series. The work shows how residual signal arising from a time-series model might be explained by exogenous information in the environment.
Implement this model and critically evaluate the claims in the paper:

First implement a simple second / third order autoregressive model on the index data and calculate the residual. The variance of this residual can be an estimate you can use as observation noise variance for the Kalman filter.

Implement a Kalman filter to estimate AR model parameters recursively. The process noise variance has to be tuned; observe if there is anything systematic about the role of this hyper-parameter. Implementing a Kalman filter is tricky, so you should first construct a synthetic time series (in which you know the correct answer to the AR model coefficients), gain some experience in working with this before applying the model to real data.

Model the error signal as a linear function of the variables considered in Mahler's paper ([4]) with a sparsity inducing regulariser (lasso) and explore their relative relevance. The residual computed from the batch solution to the AR model could be used here, if your Kalman filter did not produce stable results.

- Use the convex programming package `cvx` from `http://cvxr.com/` to implement sparse regression. You are familiar with this from the index tracking exercise.
- `S&P500` time series can be obtained at `http://finance.yahoo.com`
- Other data used in [4] can be obtained from `http://research.stlouisfed.org/fred2/`

Report

Write a report of no more than eight pages describing the work you have done.

References

[1] J. C. Hull, *Options, Futures and Other Derivatives.* Prentice Hall, 2009.

[2] P. Brandimarte, *Numerical Methods in Finance and Economics.* Wiley, 2006.

[3] J. Hutchinson, A. Lo, and T. Poggio, "A nonparametric approach to pricing and hedging derivative securities via learning networks," *The Journal of Finance*, vol. 49, no. 3, pp. 851–889, 1994.

[4] N. Mahler, "Modeling the S & P 500 index using the Kalman filter and the LagLasso," in *Machine Learning for Signal Processing, 2009. MLSP 2009. IEEE International Workshop on*, Sept 2009, pp. 1–6.

[5] R. Tibshirani, "Regression srinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B*, vol. 58, no. 1, pp. 267–288, 1996.

Mahesan Niranjan
March 2019