Quantitative Research Test

Portland House Research and Advisory

October 2019

1 Introduction

The purpose of this test is to give you a feel for the analysis of financial time series, and for us assess your ability in programming, machine learning, and statistical modelling.

Please be aware that we do not expect you to develop a highly successful strategy within a week. However we would like to see a working end-to-end pipeline, and most importantly, understand your thought process and workflow.

We ask that you submit a report detailing your approach and results for the below tasks together with your code in R/Python or any programming language that you prefer. Upload any zip and code files to a file sharing service like Dropbox or Google Drive and include the link with your PDF or MS Word report sent via email at the completion of your task.

2 Data

Data for this task can be downloaded at this link: `https://drive.google.com/open?id=1a4OXtHgcBZsqlZPG7DxcrA_m-4sef1zr`

The provided data set contains ASX stock prices from Jan 2015 to June 2018. Each csv file contains end of day (EOD) data in the following format: Ticker, Date, Open, High, Low, Close, and Volume.

- Ticker: Code referring to a particular stock
- Date: the trading day corresponding to this row
- Open / High / Low / Close: Prices in dollars corresponding to the opening / highest / lowest / closing price for specified ticker on the specified trading day
- Volume: Total number of shares traded on the specified trading day

3 Background

To make a profit trading stocks we need to be able to predict future price movement. For example, if our algorithms indicate that there is a high chance a stock will increase/decrease in price within the next day we will buy/sell that stock and hopefully make a profit.

There are many potential quantities related to price movement one could try to predict. To make things concrete for this exercise, the value we will ask you to forecast is the 1 day **future close return**. The future 1 future close return at time t, denoted $r_{t,t+1}$ is defined as:

$$r_{t,t+1} := \frac{P^c_{t+1}}{P^c_t} - 1, \tag{1}$$

where P^c_t is the close price at time t. We also define 1 day close returns at time t as $r_{t-1,t}$ - note that this quantity does not use any future information.

Your main task will be to develop a machine learning or statistical model to forecast $r_{t,t+1}$, using all the data available up to time t.

4 Tasks

4.1 Data processing

First, you are asked to write R/Python code to perform the following data processing steps:

1. Converting the EOD data into five seperate time series data frames; one each for Open, High, Low Close and Volume. In each data frame, rows should be indexed by date, and columns by ticker.

2. Create a data frame containing the future close returns as defined by Eq. (1), in the same format. Also create a data frame containing close returns.

3. Create a data frame containing the ratios of $\frac{High}{Low}$ for each ticker each day, in the same format.

> For this section, please provide your code.

4.2 Exploratory Data Analysis

Next, carry out some appropriate EDA on the above time series, with the aim of understanding key characteristics of the data, and discovering any statistical associations. You may also wish to experiment with other transformations on

the provided data which you can think of.

Some things to consider are:

- Coverage - are there missing values? If so, how could these be handled?
- The quality of the data - are there any suspect values? If so how should they be handled? can you verify these from independent sources?
- The time series structure of the data - how does this impact your analysis?

Note that there is no perfect solution to any issues you may come across, but we want to know your reasoning.

> For this section, please provide your code, along with discussion of your analysis. Any plots and tables should be properly labeled, and any extra transformations you have tried should be explained and justified.

4.3 Forecasting

We would like you to take what you have discovered in EDA and develop a model for forecasting future close returns. In addition to close returns and High/Low ratios, you can use your imagination and any feature engineering techniques you know to create more features to use in your model.
You will also need to think about how to evaluate the performance of your model. Given what you know about the distribution of the data, what metrics may be appropriate? Can you turn your model predictions into a simple trading strategy?
Please try and motivate any modelling decisions you make. Why have you chosen that particular model? How have you dealt with over-fitting? Are there any hyper-parameters that you have tuned?

> For this section please provide your code, as well as the following:
>
> - The mathematical formulation of your model, and discussion of any assumptions it may rely upon
> - Description of your modelling pipeline - how do you split the data, how do you scale the data (if at all), how do you choose hyper-parameters, etc.
> - summary of performance of your model. If you have decided to implement a simple trading strategy, describe how it works. Is it successful?

5 Final notes

If you have referred to any academic papers during the course of this task, please reference them appropriately. You don't need to list all the open source libraries you have used in your code, but if one is of particular interest feel free to mention it. Also, we will be reading your code, so please make sure it is neat enough for us to understand what you are trying to achieve (for example, an iPython notebook, or well commented script will be fine).

Thank you for taking the time to complete this task, and goodluck!