

Requirements for reports

1. The questions should be addressed in the same order they appear in the assignment. The text of the question **MUST** be retained and placed before each answer. The working language is English.
2. The answer to a particular question may take a form of a plot, formula etc followed by a brief explanation and **a conclusion**. All your conclusions **MUST** be justified numerically, i.e., by some estimates, values of a test statistic, plots, etc. For example, it is not sufficient to say that the distribution of the returns is not normal, you must refer to some numerical measures, plots etc which prove that. The answers do not need to be lengthy but, again, they **MUST** be convincing in mathematical and statistical sense, i.e., in terms of some quantitative measures. Note that I pay much attention to the conclusions, so try to make it as clear as possible.
3. Each student **MUST** use a unique data set. It is your responsibility to make sure that no one else is using the same data. **Check the list on my Google Drive and fill in your name and data you are going to use.** Here is a link to that document <https://docs.google.com/document/d/1La7Xs32Yehch5irPh6xed5jje-N2tE1pdkVK4ZhRz7s/edit?usp=sharing>.
4. When submitting your report, the **subject of your e-mail** must be **Quantitative Methods, assignment title: your name**. Otherwise, your report may get lost or not be processed properly and on time!
5. **The due date** on Black Box methods assignment is **October 14, 2017**.
6. **Late submission:** 25% off for each week after the due date.
7. The answer to a question **MUST** contain code in R or some other language which can be placed in the appendix of the report.
8. Failures to comply with the above rules may reduce your grade for the assignment.

Data sources for the assignment

1. Use `getSymbols` command of `quantmod` package to download prices for some stock or commodity (oil, gold, wheat, etc) from Federal Reserve Economic Data repository <http://research.stlouisfed.org/fred2/>, Yahoo Finance or Google Finance. You may want to try the following commands if download does not start:

```
options(download.file.method="libcurl") or options(download.file.method="wget")  
or options(download.file.method="wininet")
```

- Clearly state in your report what kind of data you are using (daily, monthly etc).
- Check for the missing data and remove the respective entries from the dataset, if any. You may use the following script as an example:

```
GOLD=getSymbols('GOLDAMGBD228NLBM', src='FRED', auto.assign=FALSE)
```

```
GOLD = na.omit(GOLD)
```

See also Section 1.3.3 of [1].

- If you did find the missing data, add a comment on that.

Assignment on Black Box methods

1. Predicting price movements. Get the log returns r_t on some stock and S&P 500 index using `getSymbols` command. Define the direction of the price movement:

$$D_t = \begin{cases} 1, & r_t > 0, \\ 0, & \text{otherwise.} \end{cases}$$

and, similarly, the market movement M_t for S&P 500 index.

- (a) Divide the data set into the training and forecast subsets. Clearly state this partition in your report.
 - (b) Fit a linear logistic regression model for $P(D_t = 1)$ using D_{t-i} , M_{t-i} , $i = 1, 2, 3$ as explanatory variables. Use only the training subset for estimation. Discuss statistical significance of the coefficients. Refine the model if needed.
 - (c) Using the model, make predictions for the forecast subset. Specify the threshold you apply. Compute the forecast error.
 - (d) Apply the linear discriminant analysis instead of log regression in items [1a-1c](#). Compute the forecast error.
 - (e) Apply the quadratic discriminant analysis instead of log regression in items [1a-1c](#). Compute the forecast error.
 - (f) Employ a 4-3-1 look-forward neural network with direct link for $P(D_t = 1)$ instead of log regression in items [1a-1c](#). Build a model for two ($i = 1, 2$) lagged variables D , M . Compute the forecast error.
 - (g) Compare the predictive power (i.e, forecast error) of methods [1b,1d,1e,1f](#).
2. Consider the trading model based on technical indicators (as discussed in class on October 4, 2017); also, see [\[2\]](#), Chapter 3:
 - (a) Choose and download the market data. Remember, the data must be unique.
 - (b) Specify the training and forecasting subsets.
 - (c) Using the template code, play with the model. Obtain the significant regressors under random forest approach.
 - (d) Try neural networks, support vector machines, regressive splines for prediction. Obtain the tabular summary of the results. Compare the predictive power.
 - (e) Also, provide the fragments of the time series with the true and predicted signals in visual or tabular form.
 3. Try to identify a pair of stocks with the common trend (as discussed in class on October 4, 2017); also, see [\[3\]](#), Chapter 6, pp. 120-132. Using this pair:
 - (a) Build a graph of spreads for prices and returns. Decide which one is “better” (in terms of mean reverting).
 - (b) Generate the signals to buy and sell for the out-of-sample spread.

References

- [1] J. Verzani, [Using R for Introductory Statistics, Second Edition](#), Chapman & Hall/CRC The R Series, Taylor & Francis, 2014.
URL <https://books.google.ru/books?id=086uAwAAQBAJ>
- [2] L. Torgo, [Data Mining with R, learning with case studies](#), Chapman and Hall/CRC, 2010.
URL <http://www.dcc.fc.up.pt/~ltorgo/DataMiningWithR>
- [3] H. Georgakopoulos, [Quantitative Trading with R: Understanding Mathematical and Computational Tools from a Quant's Perspective](#), Palgrave Macmillan, 2015.