R-evolution in Time Series Analysis Software Applied on R-omanian Capital Market

Ciprian ALEXANDRU[1*], Nicoleta CARAGEA[2], Ana - Maria DOBRE[3]

[1] Ecological University of Bucharest - Faculty of Economics, alexcipro@yahoo.com, Bucuresti 061341 Romania
[2] Ecological University of Bucharest - Faculty of Economics, nicoletacaragea@gmail.com,, Bucuresti 061341 Romania
[3] National Institute of Statistics Bucharest, dobre.anamaria@hotmail.com,, Bucuresti 050706 Romania

Abstract

Worldwide and during the last decade, R has developed in a balanced way and nowadays it represents the most powerful tool for computational statistics, data science and visualization. Millions of data scientists use R to face their most challenging problems in topics ranging from economics to engineering and genetics. In this study, R was used to compute data on stock market prices in order to build trading models and to estimate the evolution of the quantitative financial market. These models were already applied on the international capital markets.

In Romania, the quantitative modeling of capital market is available only for clients of trading brokers because the time series data are collected for the commercial purpose; in that circumstance, the statistical computing tools meet the inertia to change. This paper aims to expose a small part of the capability of R to use mix-and-match models and cutting-edge methods in statistics and quantitative modeling in order to build an alternative way to analyze capital market in Romania over the commercial threshold.

Keywords: quantmod, financial modeling, R, capital market, trading models.

1. Introduction

Forecast the financial asset prices has become a challenge in the more unpredictable and volatile world. The use of software have become very extensive in the financial field, most of the econometric models focuses on the capital market data. In this sense, *"if you don't go with R now, you will someday"*[5]. In last period of time, progress in statistics has been marked by the increasing availability of software, such as the most known and open source R system. This has the potential to continue the transformation from a set of techniques used and developed by statisticians and computer scientists to an essential system of analysis tools for a much larger community. A large area of new and improved software packages has facilitated the implementation burden for many statistical analysis methods.

At the same time, there is growing a greater recognition of the statistical computational methods using R as a powerful tool with important practical applications across a number of research areas, from economics to the social sciences and beyond of these. This trend will continue with the increasing availability of huge quantities of data and the software to analyze it. R, as a statistical application development environment, merges many forms of innovation, even if initially it belongs, as intrinsic value, of innovation itself by introducing a new concept analysis tools market data. Empowering of innovation is achieved by enabling the possibility that scientific community has to create and introduce scientific software packages, which summarize a number of functions in a particular area of research. The second form of innovation is achieved because of the potential of the scientific community to contribute to improve the existing packages by changing these functions or by adding new functions within the legal framework of open source licensing.

In this study, R was used to compute data on stock market prices in order to build trading models and to estimate the evolution of the quantitative financial market. These models were already applied on the international capital markets.

In Romania, the quantitative modeling of capital market is available only for clients of trading brokers because the time series data are collected for the commercial purpose; in that circumstance, the statistical computing tools meet the inertia to change.

[*] Corresponding author: alexcipro@gmail.com.

2. Documentation/Literature review

Finance issues have traditionally been one of the two key users of the S language, and this constituency has moved from S+ to R. In recent years, R has grown tremendously both in terms of capabilities and users, being also increasing use in Finance field. Hence, R provides an excellent platform for academic research and teaching - as well as investment research and trading.

R is used for statistical analysis, data manipulation, visualization and exciting applications in various fields like: statistics, economy, financial, business, genetics, engineering, biology and many more. One of its big advantages is the linkage with the way statisticians think and work (e.g.: keeping the track of missing values). The wide area of use - statistics, mapping, finance, forecasting, social networking, computational biology and many more – makes R a common language for all the researchers and data analysts.

The most recently used R packages for modeling the quantitative data on capital market was started to develop in 2008 and upgraded in 2013 [8]. Besides these packages, a very wide variety of functions suitable for empirical work in Finance is provided by both the basic R system (and its set of recommended core packages), and a number of other packages on the Comprehensive R Archive Network (CRAN).

Although there are many books on finance across the world, being a very new statistical tool to explore quantitative data on capital market, few deal with the statistical aspects of modern data analysis as applied to financial problems. This paper is the first step in filling this gap by addressing some of the most challenging issues facing any financial analysis.

3. Modeling financial data with R

This paper aims to expose a small part of the capability of R to use mix-and-match models and cutting-edge methods in statistics and quantitative modeling in order to build an alternative way to analyze capital market in Romania over the commercial threshold. It also supplies for academic research area an accessible approach to financial econometric models and their applications to real-world empirical research.

Econometric models can be used to predict values for next period of time. For that, many model systems in R use the same function, conveniently called *predict*. Predict is a generic function for predictions from the results of various model fitting functions. The function invokes particular methods which depend on the class of the first argument. Most prediction methods which are similar to those for linear models have an argument *newdata* specifying the first place to look for explanatory variables to be used for prediction. Some considerable attempts are made to match up the columns in *newdata* to those used for fitting, for example that they are of comparable types and that any factors have the same level set in the same order. Building a quantmod model with a given specified fitting method currently uses some function as the followings: lm, glm, loess, step, ppr, rpart, tree, random Forest, mars, polymars, lars, rq, lqs, rlm, svm, and nnet. Additional methods wrappers can be created to allow for modelling using custom functions. The only requirements are for a wrapper function to be constructed taking parameters quantmod, training.data. The function return the fitted model object and have a predict method available.

3.1. Data source

Extraction of financial time series from available Internet sources is a highly discussed topic lately, both in academia and in the specific brokers. For this there are at least two solutions. The most common are commercial solutions, but have both financial and technical barriers. Financial barriers are taxes connecting to databases or even the cost of a software environment that provides access to such databases. Technical barriers consist of dependency that we have from the software provided by the broker or the entity through which we access to these data sets. In recent years more and more available open access database and access them are designed for a variety of software environments. In this paper we have chosen to present some packages available in R, and the most used are quantmod and Quandl (Table 1, Annex1, 2).

The authors of the present study did the computation of data, but the ensuring the quality and correctness of statistical or scientific software constitutes the responsibilities of scientific software developers and scientists who provide the codes to solve a specific computational task.

Table 1. Financial Data Sources and R packages

Source	R package	Free access	Available on CRAN	Provider url
Yahoo, FRED, Oanda, Google	Quantmod	Yes	Yes	Quantmod
Quandl	Quandl	Yes	Yes	Quandl
TrueFX	TFX	Yes	Yes	TrueFX
Bloomberg	Rbbg	No	No	findata
Interactive Broker	IBrokers	No	Yes	InteractiveBrokers
Datastream	rdatastream	No	No	Datastream
Penn World Table	pwt	Yes	Yes	Penn World Table
Yahoo, FRED, Oanda	fImport	Yes	Yes	Rmetrics
ThinkNum	Thinknum	Yes	Yes	ThinkNum
DataMarket	rdatamarket	Yes	Yes	DataMarket

Source: The R Trader, "Financial Data Accessible from R – part IV", December 2013, [Online], Available: http://www.r-bloggers.com/financial-data-accessible-from-r-part-iv/ [Accessed Jan. 15, 2014]

The package quantmod has capability of downloading stock and index prices from Yahoo Finance and Google Finance and contains plotting and charting functionality. The Quandl package is a bit different in that it is tied in with the Quandl website, a source of financial data itself, as well as a portal to economic and social science data [7]. In our analysis we used version 3.0.2. R software with different packages explained adequately before running code.

3.2. Using quantmod package

The quantmod package is a *Quantitative Financial Modelling & Trading Framework* for R, designed like an environment to assist the quantitative trader in the development, testing, and deployment of statistically based trading models. The quantmod has been created to have functions which could easily use to replicate in R data modeling, so that we could access that functionality using a function with defaults and naming consistent with common usage in the finance literature.

As with any other R package, one must install the quantmod package in the usual way. As an aside, included with the installation is also the xts package for time series data, which we will discuss briefly later on.

To install the package we run the following code:

```
> install.packages("quantmod")
```

Then, we load the package:

```
> library(quantmod)
```

The traders are quite familiar with the finance sites on Yahoo and Google as sources for tracking stock, mutual fund, and exchange traded fund (ETF) prices and returns. With quantmod, we can easily load this data into R by specifying the same ticker symbol that is used in these two web sources. Also, an important data series providers is Oanda.com trader, especial for Romanian users.

```
> getSymbols("RONEUR" , src="oanda")
[1] "RONEUR"
```

The data set is returned to the R session in the form of an xts object with the name `RONEUR`. To check the contents we use the head(.) and tail(.) R functions:

```
> head(RONEUR)
           RON.EUR
2012-09-14  0.2224
2012-09-15  0.2222
2012-09-16  0.2224
2012-09-17  0.2224
2012-09-18  0.2226
2012-09-19  0.2224
```

```
> tail(RONEUR)
           RON.EUR
2014-01-10  0.2215
2014-01-11  0.2204
2014-01-12  0.2202
2014-01-13  0.2202
2014-01-14  0.2208
2014-01-15  0.2211
```

We can have access to any stock price history available on Yahoo or Google Finance, as another example, let's download Apple's stock price data series:

```
> # use single quotes and specify data source:
> getSymbols("AAPL", src = "yahoo")  # but src = "yahoo" is the default
```

Visualizing data with head(.) and tail(.) R functions:

```
> head(AAPL)
           AAPL.Open AAPL.High AAPL.Low AAPL.Close AAPL.Volume AAPL.Adjusted
2007-01-03     86.29     86.58    81.90      83.80    44225700         81.03
2007-01-04     84.05     85.95    83.82      85.66    30259300         82.83
2007-01-05     85.77     86.20    84.40      85.05    29812200         82.24
2007-01-08     85.96     86.53    85.28      85.47    28468100         82.64
2007-01-09     86.45     92.98    85.15      92.57   119617800         89.51
2007-01-10     94.75     97.80    93.45      97.00   105460000         93.79
> tail(AAPL)
           AAPL.Open AAPL.High AAPL.Low AAPL.Close AAPL.Volume AAPL.Adjusted
2014-01-08    538.81    545.56   538.69     543.46     9233200        543.46
2014-01-09    546.80    546.86   535.35     536.52     9969600        536.52
2014-01-10    539.83    540.80   531.11     532.94    10892000        532.94
2014-01-13    529.91    542.50   529.88     535.73    13517600        535.73
2014-01-14    538.22    546.73   537.66     546.39    11877200        546.39
2014-01-15    553.52    560.20   551.66     557.36    13987100        557.36
```

We can then extract the closing prices to an R vector:

```
# coerce from an xts object to a standard numerical R vector:
> AAPL_vector <- as.vector(AAPL[, "AAPL.Close"])
```

3.3. Using Datamarket from within R

The rdatamarket package is an R client for the DataMarket.com API, fetching the contents and metadata of datasets on DataMarket.com into R.

To install the package we run the following code:

```
> install.packages("rdatamarket")
```

Then, we load the package:

```
> library(rdatamarket)
```

Then, we load the package:

```
> library(rdatamarket)
```

Loading de data series from datamarket.com and plotting these data is possible with just on line of code:

```
> plot(dmseries("17tm", Country=c("Romania", "United Kingdom")))
```

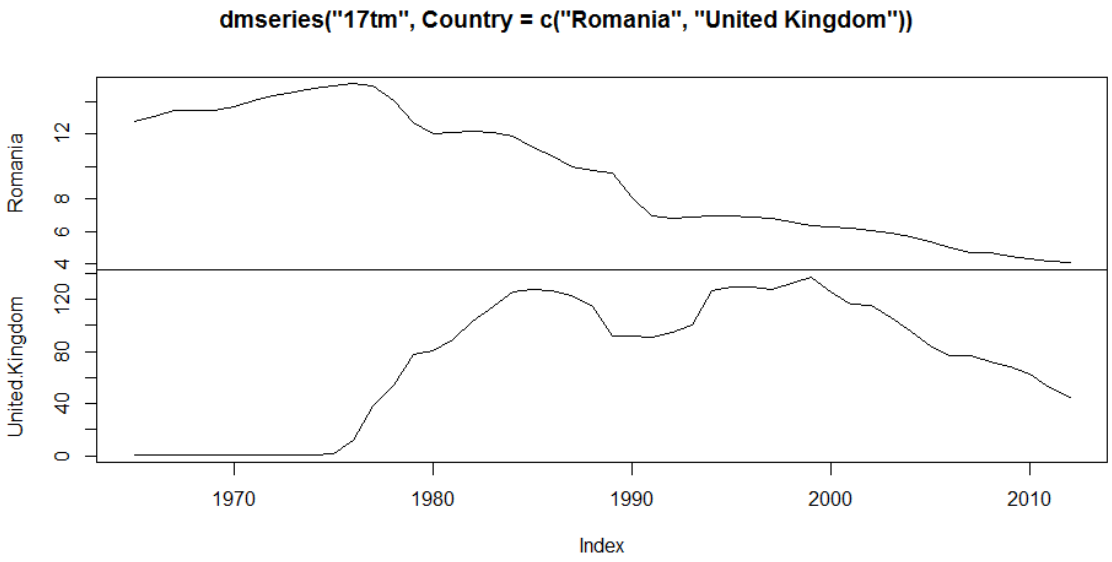

Fig. 1. Output plot for dataseries '17tm'

We can read metadata, for example, a dataset with two dimensions:

```
> p <- dminfo("http://datamarket.com/data/set/12r9/male-population-thousands")
```

Fetch the UN's population prediction for Romania and Sweden in the constant-fertility scenario.

```
> dmseries(p, 'Country or Area'=c("Romania", "Sweden"),
           Variant="Constant-fertility scenario")
```

Table 2. The UN's male population prediction for Romania and Sweden, in the constant-fertility scenario

Year	Romania	Sweden
2010	10655.359	4671.687
2015	10474.793	4828.263
2020	10241.697	4990.339
2025	9947.068	5146.432
2030	9605.094	5282.872
2035	9236.157	5404.798
2040	8849.399	5530.330
2045	8443.669	5668.217
2050	8015.484	5814.996
2055	7564.344	5954.044
2060	7096.434	6082.121
2065	6627.191	6200.868
2070	6172.537	6312.981
2075	5744.062	6416.869

Year	Romania	Sweden
2080	5350.278	6507.474
2085	4992.423	6582.797
2090	4662.182	6643.501
2095	4350.431	6688.996
2100	4053.578	6716.262

Source: datamarket.com, author's computation [Accessed Jan. 15, 2014]

This example illustrates dimension filtering and their values can be specified by their $id or their $title, to fetch the data filtered to specific values of a dimension. By default, in case we don't specify the filter, all of the dataset is fetched. In this situation be careful to some datasets, because could be very big and the DataMarket.com API may truncate extremely large responses.

4. Looking towards the future

The plans of the authors are related to continue developing in and contributing to implement the R environment in order to provide a free and open source software for data analysis in various research and academic fields in Romania and among the individual persons, investors, financial institutions and commercial and noncommercial organizations [9].

Acknowledgement

The present paper is part of the research project of Romanian R-userRs Team. The authors are thankful to their team for the great collaboration. Also, the authors of this paper would like to express the special thanks to Jeffrey Ryan, the author of the R package *Quantitative Financial Modelling Framework*. He is perhaps best known in the R community as one of the primary organizers of the annual R/Finance conference. In addition to his efforts on the R/Finance organizing committee, he has also developed a number of popular R packages for financial analysis.

References

[1] Alexandru, A.C., *"Modele de analiza a volatilitatii pietei de capital din Romania"*, Editura Mustang, Bucharest, pp. 3-212, 2013.

[2] Alexandru, A.C., *"Diversificarea instrumentelor specific pietei de capital utilizate pentru cresterea stabilitatii economic-financiare ca factor determinant al dezvoltarii sociale in spatial european"*, Editura Expert, Bucharest, pp. 7-220, 2013.

[3] Alexandru, A.C., Caragea, N., Dobre, A.M., *"Innovative methods to analyze the stock market in Romania. Studying the volatility of the Romanian stock market with the ARCH and GARCH models using the "R" software,"* Theoretical and Applied Economics, Asociatia Generala a Economistilor din Romania - AGER, vol. 0(11(588)), pp. 83-100, 2013.

[4] Caragea, N., Alexandru, A.C., Dobre, A.M., *"Bringing New Opportunities to Develop Statistical Software and Data Analysis Tools in Romania"*, The Proceedings of the VIth International Conference on Globalization and Higher Education in Economics and Business Administration, ISBN: 978-973-703-766-4, pp. 450-456, 2012.

[5] Kane, D., *"[R-sig-finance] R vs. S-PLUS",* 2004, [Online], Available: https://stat.ethz.ch/pipermail/r-sig-finance/2004q4/000186.html/[Accesed Dec, 20, 2013].

[6] Muenchen, R., *"The Popularity of Data Analysis Software"*, 2012, [Online], Available: http://r4stats.com/articles/popularity/[Accesed Dec, 20, 2013].

[7] Rickert, J., *"Quantitative Finance Applications in R"*, December 2013, [Online], Available: http://www.r-bloggers.com/quantitative-finance-applications-in-r/ [Accessed Jan. 15, 2014].

[8] Ryan, J.R., *"Quantitative Financial Modelling Framework"*, Aug, 2013, [Online], Available: http://cran.r-project.org/web/packages/quantmod/quantmod.pdf/[Accesed Dec, 20, 2013].

[9] http://www.r-project.ro/

Annexes

Annex 1

Data description

- Yahoo: Free stock quotes, up to date news, portfolio management resources, international market data, message boards, and mortgage rates that help you manage your financial life

- FRED: Download, graph, and track 149,000 economic time series from 59 sources
- Oanda: Currency information, tools, and resources for investors, businesses, and travelers
- Google: Stock market quotes, news, currency conversions & more
- Quandl: Futures prices, daily. Quandl is a search engine for numerical data. The site offers access to several million financial, economic and social datasets
- TrueFX: Tick-By-Tick Real-Time And Historical Market Rates, Clean, Aggregated, Dealer Prices
- Bloomberg: Financial news, business news, economic news, stock quotes, markets quotes, finance stocks, financial markets, stock futures, personal finance, personal finance advice, mutual funds, financial calculators, world business, small business, financial trends, forex trading, technology news, bloomberg financial news
- Interactive Broker: Interactive Brokers Group, Inc. is an online discount brokerage firm in the United States
- Datastream: Datastream Professional is a powerful tool that integrates economic research and strategy with cross asset analysis to seamlessly bring together top down and bottom up in one single, integrated application
- pwt: The Penn World Table provides purchasing power parity and national income accounts converted to international prices for 189 countries/territories for some or all of the years 1950-2010
- Thinknum: Thinknum brings financial data from a variety of useful sources together on one platform. We use this data to develop applications
- DataMarket: DataMarket brings complex and diverse data together so you can search, visualize and share data in one place and one format

Annex 2

Package Detail

- Quantmod: Specify, build, trade, and analyse quantitative financial trading strategies
- Quandl: This package interacts directly with the Quandl API to offer data in a number of formats usable in R, as well as the ability to upload and search
- TFX: Connects R to TrueFX(tm) for free streaming real-time and historical tick-by-tick market data for dealable interbank foreign exchange rates with millisecond detail
- Rbbg: Handles fetching data from the Bloomberg financial data application
- IBrokers: Provides native R access to Interactive Brokers Trader Workstation API
- rdatastream: RDatastream is a R interface to the Thomson Dataworks Entreprise SOAP API (non free), with some convenience functions for retrieving Datastream data specifically. This package requires valid credentials for this API
- pwt: The Penn World Table provides purchasing power parity and national income accounts converted to international prices for 189 countries/territories for some or all of the years 1950-2010
- fImport: Rmetrics is the premier open source software solution for teaching and training quantitative finance. fImport is the package for Economic and Financial Data Import
- Thinknum: This package interacts directly with the Thinknum API to offer data in a number of formats usable in R
- rdatamarket: Fetches data from DataMarket.com, either as timeseries in zoo form (dmseries) or as long-form data frames (dmlist). Metadata including dimension structure is fetched with dminfo, or just the dimensions with dmdims.