Telman Elchin IBRAHIMOV
Master at the International Business and Economics. Hohenheim University, Germany
E-mail: telman.ibrahimov.1998@gmail.com

ROLLİNG WİNDOW REGRESSION FOR STOCK MARKET ANALYSIS

Summary

It is historically proven that the return on stocks has always been ahead of the inflation rate in the most of developed countries. With the popularity of investing, newer methods of analyzing are appearing. While some methods are relevant for data related to population studies, the other methods fit better into econometric studies, where the financial analyses belong to. Therefore, it is noticeable to choose proper tools for analyzing financial data. This study suggests the utilization of Rolling window method for analyzing the historical data of assets to forecast the trends in theoretical basis and discusses how it can be implemented in R software.

Keywords: *Investment, finance, data analysis, r software, rolling window regression*

JEL:F02
UOT:336
DOI: https://doi.org/ 10.54414/CZGT9612

Introduction

The interest in investing refers to social problem of uncertainty about future. According to Alayli-Goebbels et al., one of them is fear that money today is of more worth than tomorrow. From economic point, the rational usage of financial resources or proper allocation can increase its amount in the future, e.g., if a person buys stock of a company that will be proactive and gain commercial success, then those stocks will be more valuable which, in turn, increases the price of the stock in market. It is believed that when one invests in stocks, s/he protects its money from inflation. If taking into consideration economically stable countries, with normal economic indicators and the stock market, in particular the US stock market, then historically, the return on stocks has always exceeded the inflation rate. Securities do not just protect money, preventing it from becoming cheaper, they also allow to significantly overtake the rise in the price level and earn more (Alayli-Goebbels et al., 2015).

Jan Jong De explains that majority of people invest with the strive for earn dividends that can be considered as passive income. In accordance with the multiple empirical studies based on this field, in the initial phase dividend streams of a firm will be small fractions if the firm recently entered to an industry or exists in the field and has an innovative product or service to launch (De, 2016).The infrastructure has made investing easier, especially for beginners. It is important to remind that various strategies could be applied when it comes to investing. One of the safest tactics is considered to develop efficient portfolio of stocks according to risk-preferences of an investor (Bernstein et al., 2013). However, before investing into any stocks, securities, cryptocurrencies or ETFs, it is important to conduct the historical analysis of asset's data in order to forecast its future trend.

Tools for analysis

There are many tools used in the financial analysis industry, but one of the most convenient application is programing language "R" for statistical analyses. This software comprises uncountable amount of packages, that can be created, added and modified by all users and then be available for all users. The availability of code chunks in the free-access websites makes it even more attracting to use R. Another issue when conducting financial analysis is the access to data sources. Although there is enormous amount of paid data source providers, some analysts prefer to use "Yahoo! Finance", online

financial media platform. It is available without registration, free of charge, and possible to download historical data of any available stock (Lee et al., 2016). When data collection process is completed, the next question of the empirical analysis is to identify the procedures to be followed.

Rolling window regression and models

Before beginning with the analysis, the historical data of chosen assets should be converted to log returns to be valid. This is explained by the assumption that the behavior of stock returns follows log normal distribution. Thus, we conclude that transformation of historical data to their log returns will allow modified data to have properties of normal distribution such as constant mean and variance (Quigley & Ramsey, 2008). Denoting the log monthly return of the asset as r_t where t indicates the return for particular time period of observation, derivation of log monthly return is calculated as:

$$r_t = \ln\left(\frac{closed\ price_{\ t}}{closed\ price_{\ t-1}}\right)$$

In this equation, "closed price $_t$" denoted the adjusted closed price of a stock at the period t.

We can now draw a conclusion of that the proper allocation of the funds can be achieved if an investor will utilize benefits of the optimal of portfolio diversification. The benefits of diversification were empirically proven, concluding on distribution of the risks for assets that were included into portfolio which, in turn, results with lower risk value than the sum of risks for individual assets (Statman & Scheid, 2008). Therefore, we include the portfolio creation into the analysis. As an addition supplementing the idea of portfolio generation is that analytical software R is equipped with packages "ROI", "PortfolioAnalytics" and "fPortfolio", that will be useful in creation different types of portfolios, including efficient, tangency and minimum variance portfolio. Afterwards, the results will be compared and proper method will be selected.

The wide variety of available econometric method and models exist and are actively implemented in researches in the field of statistics and finance. However, among of most popular methods it is better to proceed analysis with the rolling window regression analysis, which is widely used in time series analyses and practices related to financial hypotheses testing. The rolling window regression is based on so-called moving window approach, where window is the n-wide subsample that is taken away as an out-of-sample part and every time each value from 1 to n is estimated through regression repeatedly with inclusion of newly estimated value each time to the in-sample data, thus updating the values and creating forecasted estimates. These forecasted values mostly show trustworthy predictive power in historical data of financial assets due to the low variance from out-of-sample values (Zivot & Wang, 2003). Considering also the fact of the existence apt package "tsDyn" which allows us to obtain predicted values using rolling window regression through the function "predict_rolling", making a choice towards rolling window regression seems to be valid and reasonable.

However, it is important to note, that *data* must be linear object generated by either Vector Error Correction Model (VECM) or Vector autoregression (VAR) models. Therefore, it should be defined which model to use for the analysis.

Models

Before determining the suitable model to use for fitting process for further procedures with predict_rolling function, it is meaningful to explain autoregression schemes along models that are used in time series.

In essence, the forecasting cannot be exactly true as it cannot take all the details and variable in to the account by nature, so there will be some level of uncertainty in all case. The main task of applying models into the forecasting process is to achieve as lowest value of that uncertainty as possible according to all information available at the moment. The other perspectives are concerned when it comes to time series forecasting as there are only past observation on the object and there should be some connection explaining the future movements.

The first model to use with stationary time series is regression model which designed in the way where the dependent variable is linear

function of its lagged values. In other words, there is direct dependency of dependent variable on its past performance:

$$y_t = \alpha + \sum_{i=1}^{p} \beta_i y_{t-i} + \varepsilon_t$$

where α is constant value and does not depend time, $\beta_{1,\dots,p}$ are coefficients of the observation y_t in 1,…,p period steps back and ε_t is error value which has expected mean of zero. This type of regression models is known as autoregressive model and its amount of its lags are named as orders. As an example, if observation y in period t has correlation with its performance in period t-1 and t-2, then we have autoregressive model of order 2, which is indicated as AR(2):

$$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \varepsilon_t$$

Similarly, if we know this correlation coefficients, we can predict the value of the observation in the following future states as well:

$$y_{t+1} = \alpha + \beta_1 y_t + \beta_2 y_{t-1} + \varepsilon_t$$

Another model for forecasting the stationary time series data that instead of taking into consideration the lagged values of dependent observation, only considers the actual and lagged values of its error terms:

$$y_t = \alpha + \sum_{i=0}^{q} \theta_i \varepsilon_{t-q}$$

where y_t is outcome of dependent variable, α is constant parameter, θ stands for coefficient value of lagged error term and q indicated its order. Note, that this model also considers actual error in period t. Thus, coefficient for error in t, θ_0, equals to 1, which is the result of subtraction of estimated model value from actual observation value. This type of model used in time series forecasting is known as Moving Average (MA) (Enders, 2015).

The above-mentioned models are used with univariate time series, while with multivariate time series data, which is the case with several assets historical data, statisticians suggest to use vector autoregressive (VAR) and vector error correction (VECM) models. These models are only models through which predict_rolling function in R can work, therefore, these models should be reviewed as well.

VAR is, in essence, an extension of AR model allowing a user to fit multivariate times series into autoregressive model, which considers not only lagged values of the dependent variable but also lagged values of each dependent variable as independent variables. This complicated interpretation can be explained as the prediction function of a particular dependent variable through its own lags and the lags of other dependent variables that act as independent in that function. In terms of formula, the VAR(1) for 3 variables can be illustrated as:

$$y_{1,t} = \alpha_1 + \beta_{1,1} \times y_{1,t-1} + \beta_{1,2} \times y_{2,t-1} + \beta_{1,3} \times y_{3,t-1} + \varepsilon_{1,t}$$
$$y_{2,t} = \alpha_2 + \beta_{2,1} \times y_{1,t-1} + \beta_{2,2} \times y_{2,t-1} + \beta_{2,3} \times y_{3,t-1} + \varepsilon_{2,t}$$
$$y_{3,t} = \alpha_3 + \beta_{3,1} \times y_{3,t-1} + \beta_{3,2} \times y_{3,t-1} + \beta_{3,3} \times y_{3,t-1} + \varepsilon_{3,t}$$

Estimation of VAR models are usually denoted with matrices and can be processed through ordinary least squares (OLS) and then it possible to make predictions.

VECM is the extension of VAR model introduced, where it employs error correcting equations in case of co-integration. The existence of cointegration among time series variables can be checked by applying Johansen's cointegration test. If test has positive result for cointegration then VECM modelling is preferred over VAR model, however, this model is complex, responsive to outliers and requires large amount of data for analysis (Kuo, 2016).

Considering the fact that the data including price changes have already been differenced to transform data into return time series application of VECM model might lead to incorrect conclusions. As we already assume stationary time series data on returns it makes sense to proceed with VAR model for prediction in rolling window regression. This is easily done through the "lineVar" function in "tsDyn" package.

Conclusion

In conclusion, this paper outlines the importance of proper analytical tools that should be used in the financial analysis of the preferred asset(s) to invest. The field with high demand on precise way of assets analyzing have developed multiple analysis methods and tools that are actively used in stock financing.

The variety of software and data source providers makes the analysis process tougher, however, referring the wide range of functionalities and flexibility, this paper suggest to use convenient tools. In terms of analyzing software, programming language R is recommended since it has gotten various of different packages and is up-to-dated with the modern and can be equipped with emerging methods to perform data analysis. The paper also shortlists the huge amount of data source providers to the prominent "Yahoo! Finance" which is suitable both for beginners and experienced investors.

The main point is made on the mechanism of rolling window regression and its implementation in R software. In parallel, the proper choice of model fitting is described in details. Overall, it is suggested to convert the multivariate historical data into log return, fit the data into one of the models, either VAR or VECM.

References

1. Alayli-Goebbels, A. F. G., van Exel, J., Ament, A. J. H. A., de Vries, N. K., Bot, S. D. M., & Severens, J. L. (2015). Consumer willingness to invest money and time for benefits of lifestyle behaviour change: An application of the contingent valuation method. *Health Expectations*, *18*(6), 2252–2265. https://doi.org/10.1111/hex.12195

2. Bernstein, S., Lerner, J., & Schoar, A. (2013). The Investment Strategies of Sovereign Wealth Funds. *Journal of Economic Perspectives*, *27*(2), 219–238. https://doi.org/10.1257/jep.27.2.219

3. De, J. J. (2016). Time to invest? - Money matters. *The Dairy Mail*, *23*(4), 73–75. https://doi.org/10.10520/EJC187652

4. Enders, W. (2015). Applied econometric time series fourth edition. *New York (US): University of Alabama*.

5. Kuo, C.-Y. (2016). Does the vector error correction model perform better than others in forecasting stock price? An application of residual income valuation theory. *Economic Modelling*, *52*, 772–789. https://doi.org/10.1016/j.econmod.2015.10.016

6. Lee, C.-F., Lee, J., Chang, J.-R., & Tai, T. (2016). Data Collection, Presentation, and Yahoo Finance. In *Essentials of Excel, Excel VBA, SAS and Minitab for Statistical and Financial Analyses* (pp. 13–45). Springer.

7. Quigley, L., & Ramsey, D. (2008). Statistical analysis of the log returns of financial assets. *Financial Mathematic, University of Limerick*, *32*.

8. Rao, B. B. (1997). *Cointegration for the applied economist*. Allied Publishers.

9. Statman, M., & Scheid, J. (2008). Correlation, Return Gaps, and the Benefits of Diversification. *The Journal of Portfolio Management*, *34*(3), 132–139. https://doi.org/10.3905/jpm.2008.706250

10. Zivot, E., & Wang, J. (2003). Rolling analysis of time series. In *Modeling Financial Time Series with S-Plus®* (pp. 299–346). Springer.

Тельман Эльчин оглы ИБРАГИМОВ

Магистр по международному бизнесу и экономике. Университет Hohenheim, Германия.

E-mail: telman.ibrahimov.1998@gmail.com

Резюме

Исторически доказано, что доходность акций всегда опережала уровень инфляции в большинстве развитых стран. С ростом популярности инвестирования появляются новые методы анализа. Хотя некоторые методы применимы для данных, связанных с демографическими исследованиями, другие методы лучше подходят для эконометрических исследований, к которым относится финансовый анализ. Поэтому важно выбирать

правильные инструменты для анализа финансовых данных. В этом исследовании предлагается использовать метод скользящего окна для анализа исторических данных об активах для прогнозирования тенденций в теоретической основе и обсуждается, как его можно реализовать в программном обеспечении R.

Ключевые слова: *Инвестиции, финансы, анализ данных, программное обеспечение r, регрессия скользящего окна.*

Telman Elçin oğlu İBRAHİMOV
Beynəlxalq Biznes və İqtisadiyyat üzrə magistr Hohenheim Universiteti, Almaniya
E-mail: telman.ibrahimov.1998@gmail.com

Xülasə

Tarixən sübut edilmişdir ki, əksər inkişaf etmiş ölkələrdə fond gəlirləri həmişə inflyasiyanı üstələmişdir. İnvestisiyaların artan populyarlığı ilə yeni təhlil üsulları ortaya çıxır. Bəzi metodlar demoqrafik tədqiqatlarla bağlı məlumatlar üçün uyğun olsa da, digər üsullar maliyyə təhlili daxil olmaqla ekonometrik tədqiqatlar üçün daha uyğundur. Buna görə də maliyyə məlumatlarını təhlil etmək üçün düzgün alətləri seçmək vacibdir. Bu tədqiqat nəzəri çərçivədə trendin proqnozlaşdırılması üçün tarixi aktiv məlumatlarının təhlili üçün sürüşmə pəncərə metodunun istifadəsini təklif edir və onun R proqram təminatında necə həyata keçirilə biləcəyini müzakirə edir.

Açar sözlər: *İnvestisiya, maliyyə, data analiz, R proqramı, hərəkətli pəncərə reqresiyası.*