УДК 004.9

TESTING K-VALUE IN K-FOLD CROSS VALIDATION OF FORECASTING MODELS FOR TIME SERIES ANALYSIS OF G-SPREADS OF TOP-QUALITY RUB BONDS

Elena Latysh[1], Oleksiy Koshulko[2]

[1]*Russian Presidential Academy of National Economy and Public Administration, Vernadskogo prospect 82, Moscow, Russia.*

[2]*Glushkov Institute of Cybernetics, NASU, acad. Glushkov avenue 40, Kiev, Ukraine*

elenalatysh@gmail.com, koshulko@gmail.com

Метою роботи є пошук значень величини k для оцінювання адекватності моделей на основі k-згорток у конкретній задачі, пов'язаній з прогнозуванням прибутковості облігацій для інвестиційної стратегії пенсійних фондів. Даний короткий опис системи GS та результати експериментів.

Ключові слова: індуктивне моделювання, GMDH Shell, рубльові облігації

The paper aims to find the best k-values for k-fold based model validity in a particular problem related to the forecasting of bond yields for pension funds investment strategy. The GS system and results of experiments are shortly described.

Keywords: *Inductive modeling, GMDH Shell, Rub bonds*

Целью работы является нахождение величины k для оценки адекватности моделей на основе k-сверток в конкретной задаче, связанной с прогнозированим доходности облигаций для инвестиционной стратегии пенсионных фондов. Дано краткое описание системы GS и результаты экспериментов.

Ключевые слова: индуктивное моделирование, GMDH Shell, рублевые облигации

1 Introduction

The effectiveness of pension funds investment became one of the most important questions of the Russian financial institutions. So, the focus on developing a strategy of active portfolio management of RUB bonds is actual direction for the research. Efficiency of returns on investments forecasting can be improved by applying the methods that take into account hidden information in experimental data. GMDH provides such an opportunity.

Group Method of Data Handling (GMDH) is a family of inductive algorithms for computer-based mathematical modeling of multi-parametric datasets that features fully automatic structural and parametric optimization of models. GMDH is an original method for solving problems of structural and parametric identification under conditions of uncertainty. The method was originated in 1968 by acad. Olexiy Ivakhnenko at the Glushkov Institute of Cybernetics in Kiev and nowadays his colleagues and students develop it under the supervision of Volodymyr Stepashko (Ukraine) [1, 2]. GMDH repeatedly proved its effectiveness including the examples of economic indicators forecasting (inflation in USA)

GMDH selects the model of optimal complexity and such a selection depends on the form of external criterion realization. *K*-fold cross validation is one of such criteria and this question has already been under consideration [3]. In the paper we experimentally study *k*-fold cross validation with respect to a concrete problem: forecasting of bond yields for pension funds investment strategy

2 Source data

2.1 Possibilities of GMDH Shell

Experiments with GMDH were carried out using the GMDH Shell (GS) program – a predictive modeling tool that produces mathematical models and predictions. It was developed by Geos company under the leadership of Olexiy Koshulko (Ukraine). GS contains a number of modifications of the original GMDH [4]

There are two learning algorithms available in GS:

- Combinatorial GMDH
- GMDH-type neural networks

Learning algorithm is a procedure that controls model optimization process. In the modern theory of predictive modeling it is well known that the model should provide a trade-off between simplicity and accuracy. GMDH Shell matches this goal using learning algorithms of the Group Method of Data Handling.

The experimental models in the research were linear combinations of values at the previous time intervals, taking into account the mutual ties, lags, etc. The model optimality is understood as the best form of the model in a given class of predictive models. The results of the forecast are the basis for the choice of NPF investment strategies.

2.2 Original source of data

The following input parameters were used to conduct the forecasting model:

- Values of G-Spreads of 28 ruble bonds of Russian banks with the highest credit risk quality.
- OFZ yield values.

Input parameters – daily values of 12 months period. The daily average values are calculated on the basis of data from MICEX Stock Exchange (MICEX, RTS); the bond yield was calculated in accordance with generally accepted methodology of G-bonds.

G-Spread is a spread between the bond yield and government curve. A government curve used as a G-Curve (daily G-Curve data, source: Cbonds.ru), which is generally

accepted base curve for a financial market participants. G-Spread is extremely important bond parameter for investors, so the prediction of its values is the basis of investment decisions making: to buy or sale bond.

Forecast horizon fixed at 1, 5, and 10 values which corresponds to maximum two-week forecast.

3 Experiments

The experiment was carried out using k-fold cross-validation option. The original sample was randomly partitioned into k subsamples. A single subsample was taken as the validation data for testing model, and the other k−1 subsamples were used as training data. The cross-validation process was repeated k times with each of the k subsamples used exactly once as the validation data. Then the k results from the folds then can produce a single estimation. The advantage of this method over repeated random sub-sampling is that all observations are used for both training and validation, and each observation is used for validation exactly once.

The experiment was carried out using root-mean-square error (RMSE) validation criterion, which selects models with the lowest RMSE on the testing sample.

The experiments were carried out for different values of k (from 2 to 10). The optimal time series forecasting model was selected by minimum RMSE Critarion Value. As we can see on Figure 1 forecasting model is optimal for k=3

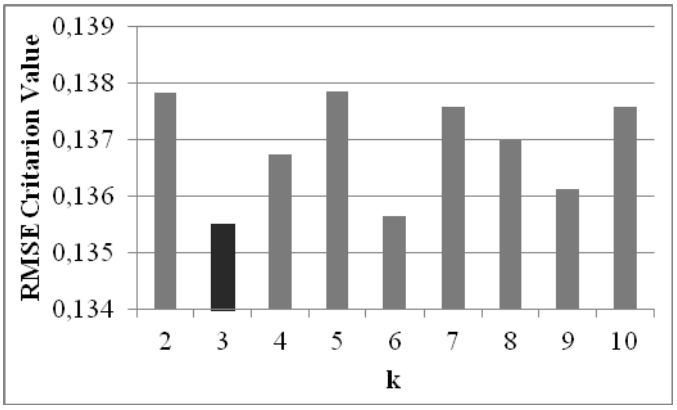

Fig. 1. RMSE Criterion Value for different k.

4 Conclusions

The optimal value k=3 was obtained in our experiments. The optimal time series forecasting model for predicting of RUB bond yield dynamics was selected.

Research results can be displayed for the following bond (RusAgroBank, 07):

Fig. 2. RusAgroBank Bond forecast (k=3).

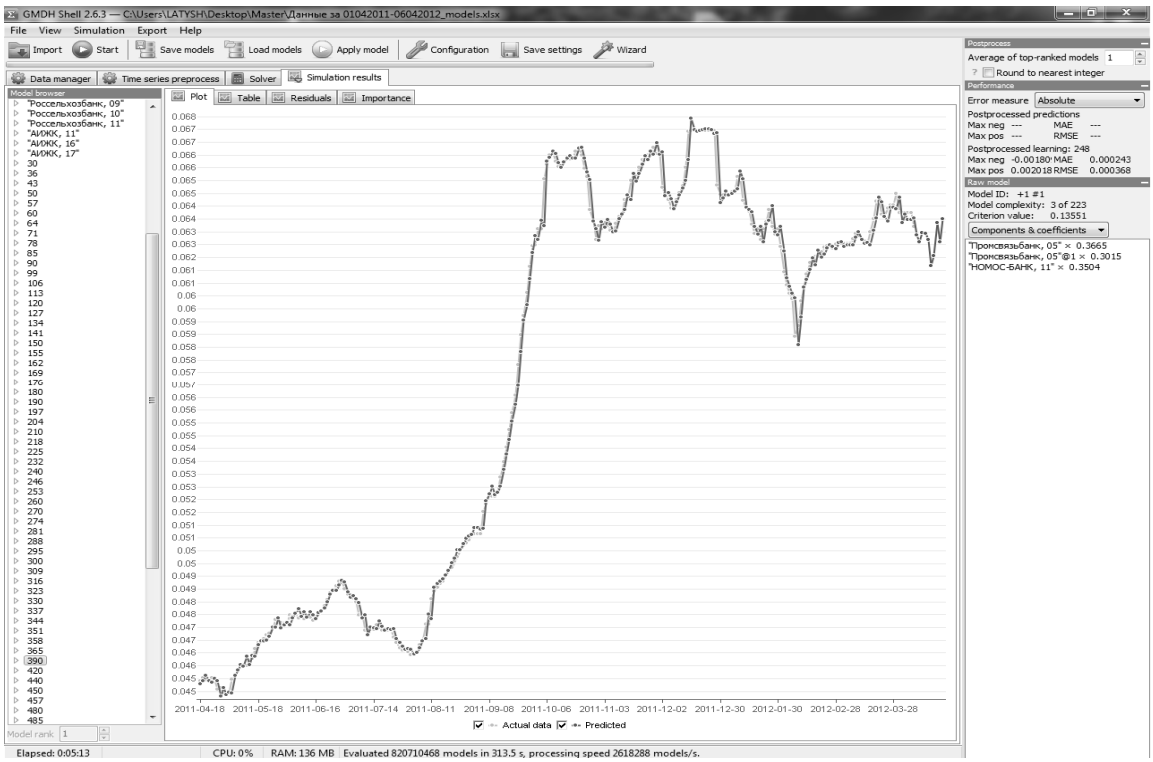

Fig. 3. OFZ yield forecast for RusAgroBank Bond (k=3).

In accordance with the optimal model in 10 forecast periods the RusAgroBank bond G-Spread is expected to expand (Figure 2) together with constant OFZ yield with the appropriate duration (Figure 3). It can be identified as a bond price reduction in 2 weeks. According to the model the recommendation is to sell the RusAgroBank bond. The same reasoning can be applied to each bond used in the model. The investment strategy of NPF can be based on the result of the optimal model forecast according to table 1.

Tab.1.

Decision making process depending on forecasting model results

	OFZ yield increase	*OFZ yield constancy*	*OFZ yield decrease*
G-Spread narrowing	Sell OFZ	Buy Bond	Buy Bond/ Buy OFZ
G-Spread constancy	Sell OFZ	-	Buy OFZ
G-Spread expanding	Sell Bond/ Sell OFZ	Sell Bond	Buy OFZ

References

[1] Ivakhnenko A.: Polynomial theory of complex systems. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-1(4), 1971, pp. 364-378

[2] Ivakhnenko, A., Stepashko, V.: Noise Immunity of Modeling. *Kiev, Naukova Dumka Publ.*1985 (rus.)

[3] Koshulko, O., Koshulko, G.: Validation Strategy Selection in Combinatorial and Multilayered Iterative GMDH Algorithms. In: *Proc. Intern. Workshop on Inductive Modeling (IWIM-20011)* Kyev, 2011, pp. 51-54

[4] Koshulko, O., Koshulko, A.: Adaptive parallel implementation of the Combinatorial GMDH algorithm. In: *Proc. Intern. Workshop on Inductive Modeling (IWIM-2007)* Prague, 2007, pp. 71-74