　　　　　　　　　　　　　　　　　　　　　　　Open Access

The method of residual-based bootstrap averaging of the forecast ensemble

Vera Ivanyuk*

*Correspondence:
ivaver6@gmail.com

Department of Data Analysis
and Machine Learning, Financial
University, 49 Leningradsky
Prospekt, Moscow, Russia 125993

Abstract

This paper presents an optimization approach—residual-based bootstrap averaging (RBBA)—for different types of forecast ensembles. Unlike traditional residual-mean-square-error-based ensemble forecast averaging approaches, the RBBA method attempts to find optimal forecast weights in an ensemble and allows for their combination into the most effective additive forecast. In the RBBA method, all the different types of forecasts obtain the optimal weights for ensemble residuals that are statistically optimal in terms of the fitness function of the residuals. Empirical studies have been conducted to demonstrate why and how the RBBA method works. The experimental results based on the real-world time series of contemporary stock exchanges show that the RBBA method can produce ensemble forecasts with good generalization ability.

Keywords: Forecast ensembles, Time series, Artificial neural networks

Introduction

The stock exchange data forecasting problem is very large and complex when traditional statistical methods are used alone. There are several examples of incorrect forecasts from the real world that are made by even the best specialists and forecasting business-consulting centers. Even the success of artificial neural networks (ANNs) and ANN ensembles in forecasting (Liu and Yao 1999) is insufficient to avoid typical ANN problems, such as too-deep-to-learn, too-wide-to-forget, or overfitting, whereby surplus memory may be mistakenly regarded as a good learning result. Different types of assets, as well as ample dependencies and external factors, make ANN-based stock exchange forecasting a very difficult task. In practice, it is possible to obtain reliable but not sufficiently accurate predictions via simple standard methods such as SARIMAX, single-layer neural networks, wavelets, exponential smoothing, asymptotic linear regression, and so on.

This paper deals exclusively with random processes that occur on the stock exchange and are in one way or another implicitly related to human behavior. The random processes under consideration differ from the high-frequency ergodic random processes occurring in nature.

It would be naive to believe that random stock exchange factors related mainly to human behavior fully correspond to the classical concepts of the statistics of

stochastic processes. If there were such a correspondence, most economists would have been able to unequivocally predict both the beginning of the Arab Spring and the crisis of 2008, as well as the beginning of the energy crisis of 2022, a few months before these events occurresd.

This problem primarily implies the adaptation of predictive methods based on the actual situation. Unfortunately, the currently available methods of expert forecasting (references to EIA forecasts from 2017 to 2021) as well as the classical methods of statistical forecasting proposed more than 100 years ago by Fienberg and Lazar (2001) are rather weak. The objective of this study is to develop a statistical evaluation method vis-à-vis the performance of a forecast that provides a more accurate result than the currently known predictive ensemble methods.

In this case, the major task is to combine forecasts such that the residuals exclusively contain the random walk component or express it most accurately (Hashem et al. 1994). In most cases, any optimal ensemble gives better forecasting results than individual methods (Kuncheva and Whitaker 2003). Meanwhile, it is almost generally accepted that an increase in the differences in extrapolation methods enhances the overall quality of forecasts (Gashler et al. 2008). Considering both traditional methods of ensemble optimization, such as the Bayes optimal classifier (Mitchell 1997), bootstrap aggregating (Salman et al. 2021), boosting (Emer 2018), and advanced ensemble methods based on statistical estimates of the forecast correlation (Liu and Yao 1999), the author reflects on a tendency that is common to all methods and consists of using the residual mean square error (RMSE) or its variations as a generally accepted geometric measure of forecast error. Paying attention to its properties, such as unidimensionality and asymmetry of the estimate, as a function of the area of the deviation bounds, the author has conducted a study that allows for the development of a multidimensional geometric evaluation function of the ensemble residuals that is easily applicable to bootstrap averaging as a penalty function and shows quite interesting results, especially for large samples.

Considering the works of contemporary authors, Kou et al. (2021) developed a model for predicting bankruptcies of small and medium-sized enterprises. This model uses transactional data and two-stage multi-objective feature selection. Wen et al. (2019) revealed a correlation between investor interest in an asset and asset quality. The better the asset, the higher the investor's interest in it. Conversely, if the investor's interest in the asset decreases, the asset soon collapses. Li et al. (2022) proposed a new clustering approach. They improved the multidimensional K-means algorithm, which allowed for better data clustering. This approach helps to reduce the number of clusters and obtain a more accurate image.

The remainder of this paper is organized as follows: Sect. "Multidimensional evaluation of residuals: idea and implementation" describes the idea and implementation of a multidimensional function of the geometric evaluation of the ensemble residual quality and the residual-based bootstrap averaging (RBBA) method itself. Section "Experimental verification of the RBBA method" describes the conditions and procedures for experimenting with a comparative analysis of the RBBA and traditional predictive ensemble methods. Section "The results of the experiment" presents the results of the real stock asset forecasting experiment, compares them with standard Bayesian optimization,

AdaBoost, and bootstrap aggregating methods, and provides some discussions. Finally, Sect. "Discussion" concludes with a summary of the paper and a few remarks.

Multidimensional evaluation of residuals: idea and implementation

Computation age and artificial neural networks

It is usually assumed that RMSE minimization is a fairly good criterion for network training or prediction optimization, but the practical application of this method does not always present adequate results. Having agreed that the residuals of the approximation of the method should have minimum error, we nevertheless do not consider the basic statistical idea of the random component (Wetherill 1981). Often, an overfitted neural network or an idealized predictive method, being sufficiently trained, optimized, and showing good results on homogeneous, standardly distributed data, still turns out to be ineffective for practical application. With the stock market, this means that, in fact, the incoming data are not only heterogeneous but also have a purposeful tendency to deviate from statistically sound forecasts because a high-quality forecast that becomes a matter of common knowledge is actually a factor that minimizes the agent's profit. Thus, the use of standard methods of predictive error estimation forces the forecaster to deliberately expand the forecast corridor to avoid real estimates beyond theoretical boundaries. Of course, the least squares method, invented by Gauss more than 200 years ago (Stigler 1981), is currently one of the most commonly used optimization criteria for approximating dependencies. This method is extremely simple and requires minimal computing resources. However, it does not guarantee that it will separate the main functions and dependencies from a random component; it only finds its most plausible placement. As a rule, the evaluation of approximation residuals is more often used as a verification method rather than an optimizing one; traditionally, in the pre-computer and early computer era (primary and secondary information age), it was considered highly resource-intensive. Since 2008, the computing power of nonspecialized private systems has grown so much that we have entered a new computing era (tertiary information age) (Hilbert 2020). Modern computing resources have become widely available and extremely cheap over the past half-century, and the emergence of multi-core video processors (Corana 2015), method of massive error backpropagation (Goodfellow et al. 2016), and software libraries for the Levenberg–Marquardt algorithm (Transtrum and Sethna 2012) have allowed the widespread use of neural networks and computational ensembles as universal approximating and predictive mathematical tools. These three tools, mutually reinforcing one another, have created the conditions for a technological breakthrough in the field of processing and analyzing large amounts of data in small laboratories. The availability of supercomputing capacities has ceased to play a significant role because the computing power of conventional desktop computing systems has reached hundreds of teraflops.

Nearly every modern forecasting method uses the squared deviation or its modifications in one form or another, such as the fitness penalty function and the Levenberg–Marquardt algorithm, sometimes combined with stochastic functions, as a method of finding the optimum. Although the algorithm for finding a local or global error minimum solely affects the speed of achieving the result, the fitness penalty function is responsible for its quality. The idea of improving the quality of the forecast by replacing the optimization criterion

appeared during the actual forecasting of stock data, when many forecasts of real data turned out to be of insufficient quality despite the low standard error and excellent approximation quality.

The idea of a new quality criterion of residuals

The main criteria for the quality of random residuals, regardless of whether they appeared as a result of multiplicative (Bessel distribution), additive (Gaussian distribution), or mixed effects of random factors, are the actual characteristics reflecting their association with a set of random processes. Such characteristics include, first, the mutual independence of residuals of the time series, the absence of autocorrelation, and plausibility expressed in proximity to the normal distribution (symmetry, pronounced modality, moderate excessivity). Such a synthetic criterion may be multidimensional and expressed geometrically, similar to the RMSE, in terms of volume or hypervolume corresponding to the deviation from the set of ideal values of statistical tests of residuals corresponding to zero values.

Mathematical implementation of the penalty function

The following criteria were selected as components of the residual quality measure: the Bienaymé turning point test (Kendall 1973), as a characteristic of the mutual dependence of time series elements; the Durbin–Watson statistic (Durbin and Watson 1950), as a characteristic of the presence of autocorrelation; and the Shapiro–Wilk test (Shapiro and Wilk 1965), as a measure of deviation from normality.

Modified turning point test

A test first described by Bienaymé (1874) and further considered in detail by Kendall (1973) characterizes the presence of connectivity of time series elements. In most of the literature (Brockwell and Davis 2002), as well as in the implementation of statistical software packages, such as R (the turningpoint.test function), counting turning points is a rather slow procedure because it uses two comparison operations and, accordingly, branches for each point. We have developed and first introduced an equation that allows it to speed up this procedure because it does not use comparisons and branches but is based on the properties of the modulus of a number and the sign function (1). This equation contains a single slow operation (division) performed at the end of the calculations and can be optimized for data flow processing. Typically, the total number of turning points of a series, calculated as (1), is estimated only by the left boundary of the criterion, checking for the presence of linear dependencies between the elements of the series. However, the right boundary also characterizes the presence of a connection, but unlike the left boundary, it is periodic. Thus, the necessary components normalized to the optimum corresponding to the zero value can be expressed as (2).

$$T = \sum_{i=2}^{n-1} \frac{sgn(|x_t| - |x_{t-1}|) \cdot sgn(|x_t| - |x_{t+1}|) + 1}{2}$$
$$\cdot \sum_{i=2}^{n-1} \frac{sgn(|x_t| - |x_{t-1}|) \cdot sgn(|x_t| - |x_{t+1}|)}{2}. \tag{1}$$

$$T_p = \frac{\left| T - \frac{2n-4}{3} \right|}{z_\alpha \sqrt{\frac{16n-29}{90}}}, \tag{2}$$

where respectively, T is the actual number of turning points, n is the sample size of the approximation residuals, $sgn(x)$ is the sign function of the number [Kronecker function (Rich and Jeffrey 1996)], x_t are valuefunction of the number of the time series of the approximation residuals, t denotes the number of sample elements corresponding to the time point of the series, T_p is the penalty function component, and z_α is the width of the boundaries of critical values (Z-score).

Modified Durbin–Watson statistic: because the optimum point for this criterion (corresponding to 2) and the extreme values of the upper and lower bounds (corresponding to 0 and 4) are known, it is sufficient to use Eq. 3 to calculate the normalized estimate of the absence of autocorrelation.

$$D_p = \frac{\left| \frac{\sum_{t=2}^{n}(x_t - x_{t-1})^2}{\sum_{i=2}^{n} x_t^2} - 2 \right|}{2}, \tag{3}$$

where respectively, D_p is the penalty function component, n is the sample size of the approximation residuals, x_t are the values of the time series of the approximation residuals, and t is the number of sample elements corresponding to the time points of the series.

Similar to the previous component, this modified criterion is normalized to one, and its optimum point corresponds to zero.

Modified Shapiro–Wilk test

This criterion which determines the plausibility of the normality of the residual distribution has limited statistical tables that do not allow estimating time series of more than 50 elements. To expand its applicability, we can use the Kazakevičius approximation (Kazakevičius 1988), expressed by Eqs. (4)–(6), which allows working with any sample size, normalized to one with an optimum at zero.

$$z_t = \frac{n - 2t + 1}{n - 0.5}, \tag{4}$$

$$a_t = \left(\frac{0.899}{(n - 2.4)^{0.4162}} - 0.02 \right) \cdot \left[z_t + \frac{1483}{(3 - z_t)^{10.845}} + \frac{7161 \cdot 10^{-8}}{(1.1 - z_t)^{8.26}} \right], \tag{5}$$

$$W_p = \left(1 - \frac{0.6695}{n^{0.6518}} \right) \cdot \frac{\frac{1}{n-1} \sum_{t=1}^{n} (x_t - \bar{x})^2}{\frac{1}{n-1} \left[\sum_{t=1}^{[n/2]} a_t (x_{n-t+1} - x_t) \right]^2}, \tag{6}$$

where respectively, z_t is the approximated tabular value used to calculate the coefficients of a_t, n is the sample size of the approximation residuals, t is the number of sample elements corresponding to the time point of the series, a_t is the approximated tabular value

for calculating the value of W_p, W_p is the penalty function component, x_t is the values of the time series of the approximation residuals, and $\bar{x}$ denotes the arithmetic mean of x_t.

Despite the apparent complexity of Kazakevičius' equations, unlike the original Shapiro–Wilk test, they not only allow the evaluation of a sample of any volume, but are also well suited for parallelizing calculations on multiprocessor systems. In addition, it should be noted that the coefficients z_t and a_t are essentially tabular values that are calculated once and appear to be common to all estimated time series residuals.

Penalty function: as a result, having characteristics of the series of residuals as mutual independence, characterized by the value T_p, the absence of autocorrelation, characterized by the value D_p, and the proximity of the distribution to the normal, characterized as W_p, considering that all three parameters have critical values normalized to unity with an optimum point at zero, we can create a function characterizing all three parameters as a volume normalized in the interval [0; 1] with an optimum at the origin point. The final function is given by Eq. (7).

$$P_p = \sqrt{T_p\, D_p\, W_p}. \tag{7}$$

In the future, we will consider RBBA to minimize the penalty function P_p of the time series residuals by selecting the optimal values of the ensemble of variable coefficients of the approximated series by any gradient or stochastic method.

Experimental verification of the RBBA method

For the experimental verification of the method, a statistical assessment of the quality of additive ensemble forecasts formed by the following ensemble methods was applied: Bayes optimal classifier (Mitchell 1997), bootstrap aggregating (Breiman 1996), Adaptive Boosting (Hastie et al. 2009), and particularly, RBBA.

Ensemble components: The following types of forecasts were used as the main components of the additive predictive ensemble: classical linear, classical indicative, asymptotic linear, asymptotic indicative, classical wavelet, combined neural, and classical perceptron-based.

The classical and asymptotic linear forecasts are described by Eqs. (8) and (9):

$$F(t)_{lin} = at + b, \tag{8}$$

$$F(t)_{asl} = F(t)_{lin} + (F(n)_{lin} - x_n)e^{-c(t-n)\frac{e}{\tau}}, \tag{9}$$

where respectively, $F(t)_{lin}$ denotes the value of the linear forecast, a is the coefficient of the linear equation selected by the optimization method, t denotes the number of sample elements corresponding to the time point of the series, b is the coefficient of the linear equation selected by the optimization method, $F(t)_{asl}$ is the asymptotic linear forecast value, n is the sample size of the time series elements, e is the Euler number (McCartin 2006), c is the asymptotic coefficient selected by the optimization method, and τ is the forecast lead time.

The classical and asymptotic indicative forecasts are described by Eqs. (10) and (11):

$$F(t)_e = H(a)\left(x_{max} - x_{min}e^{-bt}\right) + H(-a)\left(x_{min} + x_{max}e^{-bt}\right), \tag{10}$$

$$F(t)_{ae} = F(t)_e + (F(n)_e - x_n)e^{-c(t-n)\frac{e}{\tau}}, \tag{11}$$

where respectively: $F(t)_e$ is the indicative forecast value, H is the Heaviside function (Zhang and Zhou 2020), a is the coefficient of the linear equation selected by the optimization method, x_{max} is the maximum element of the time series, x_{min} is the minimum element of the time series, e is the Euler number (McCartin 2006), b is the coefficient of the linear equation selected by the optimization method, t denotes the number of sample elements corresponding to the time point of the series, $F(t)_{ae}$ is the asymptotic indicative forecast value, n is the sample size of the time series elements, c is the asymptotic coefficient selected by the optimization method, and τ is the forecast lead time.
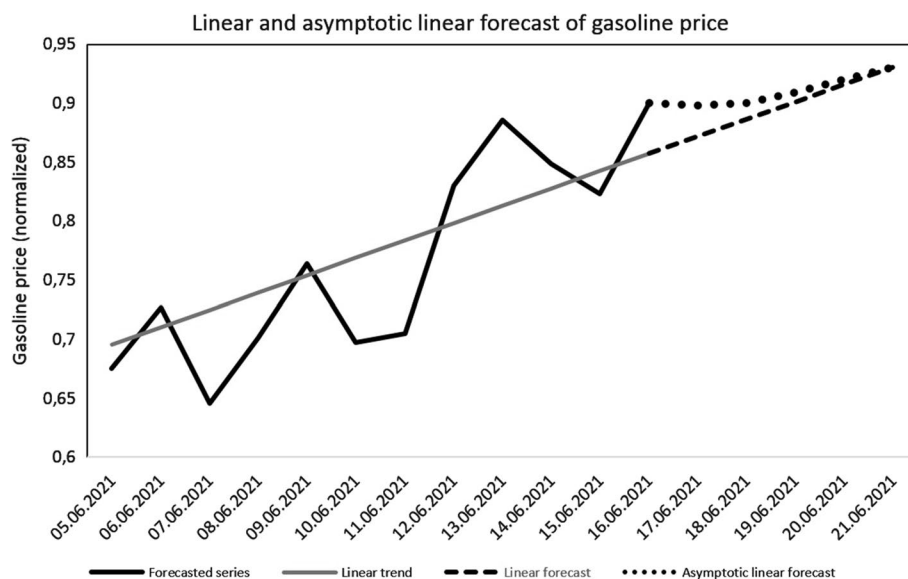
Because asymptotic forecasts are non-standard methods (Figs. 1 and 2), we suggest a graphic presentation of the difference between these forecasts and the existing classical methods, which allows for enhancing the prediction quality as a result of the asymptotic expansion of the forecast values.

The wavelet forecast is formed by the method of sequential decomposition of a time series by a set of wavelets similar to Eq. (12), all the wavelets are combined into a common additive forecast, according to Eq. (13).

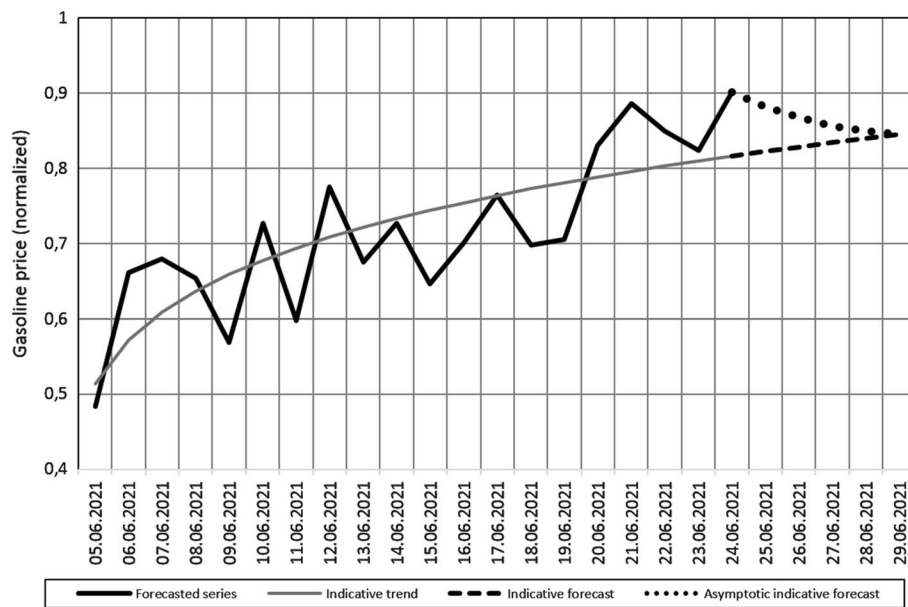$$\psi_i(t) = a_i(\sin b_i(t + d_i))e^{-|c_i(t+d_i)|}, \tag{12}$$

$$F(t)_{wav} = \sum_{i=1}^{m} \psi_i(t), \tag{13}$$

where respectively, $\psi_i(t)$ is the value of the wavelet, a is the amplitude of the wavelet selected by the optimization method, i is the number of wavelets obtained during

Fig. 1 Asymptotic linear forecast

Fig. 2 Indicative and asymptotic indicative forecast of gasoline price

decomposition, b is the Wavelet compression coefficient selected by the optimization method, t is the number of sample elements of the time series corresponding to the time point of the series, d—the wavelet shift coefficient selected by the optimization method, e is the Euler number (McCartin 2006), c is the Wavelet decay coefficient selected by the optimization method, $F(t)_{wav}$ is the wavelet forecast value, and m is the number of wavelets obtained during decomposition of the series.

A perceptron-based neural forecast, which reflects the impact of external correlating factors, is formed by a single-layered nonrecurrent neural network with multiple inputs receiving data on the values of influencing external factors. Such a forecast is calculated according to Eq. (14):

$$F(t)_{ann} = \varsigma k_0 \left(\varsigma k_1 \sum_{i=1}^{i=m} \varsigma k_i \left(\sum_{j=i}^{j=n} \varsigma \left(k_{ij} x_{ij}(t - \tau) \right) \right) + \dots \right), \tag{14}$$

where respectively, $F(t)_{ann}$ is the value of the neural network forecast, ς is the Verhulst logistic function (Verhulst 1838), m is the unit depth of the neural network, k is the coefficients of the receptive field selected by the network training method, n is the unit width of the neural network, x_{ij} is the input value of the neural network, t is the number of sample elements of the time series corresponding to the time point of the series, and τ is the forecast lead time.

Residual-based optimization function.

To perform optimization using gradient methods in an empirical study, a four-dimensional criterion of the residual quality was chosen, combining both the classical least-squares method and the proposed method for evaluating the quality of residuals corresponding to Eq. (15)

$$P_p = \sqrt[4]{T_p \cdot D_p \cdot W_p \cdot RMSE_p},\tag{15}$$

The results of the experiment

General equation of the forecast and optimization of confidence coefficients based on the forecasts described above, an additive equation of the predictive ensemble (16) is compiled:

$$F(t)_{ens} = k_l F(t)_l + k_{al} F(t)_{al} + k_e F(t)_e + k_{ae} F(t)_{ae} + k_w F(t)_w + k_n F(t)_n.\tag{16}$$

where respectively, $F(t)_{ens}$ is the ensemble forecast value, k is the coefficient of receptive confidence selected using the ensemble averaging method, and $F(t)$ is the corresponding forecast described above.

The main optimization task for this type of forecast is to determine the confidence coefficients, which is the concept of ensemble optimization (ensemble averaging). Notably, it is just the method of determining these coefficients that constitutes the main difference between the basic approaches of the evaluated ensemble methods. Therefore, for the Bayes optimal classifier method, the probability of a correct forecast for each of the ensemble components determines the confidence coefficients; bootstrap aggregating is based on discarding the worst forecasts characterizing certain properties of the series already used in forecasts with better results; Adaptive Boosting, despite the pronounced tendency to overfitting, relies on the idea of gradual improvement of confidence coefficients based on a continuous reassessment of the squared forecast deviations; and RBBA, combining the ideas of Adaptive Boosting and bootstrap aggregating, implies a gradual sequential selection of confidence coefficients based on the optimization of comprehensive quality estimates of predictive residuals of a set of forecasts.

To assess the quality of the forecasts, a random sample of real stock exchange data was used. The main characteristics of the forecast quality are the estimates of the standard deviation (17), confidence interval (18), and total percentage of forecasting errors with the corresponding lead time. The evaluation results are averaged arithmetically.

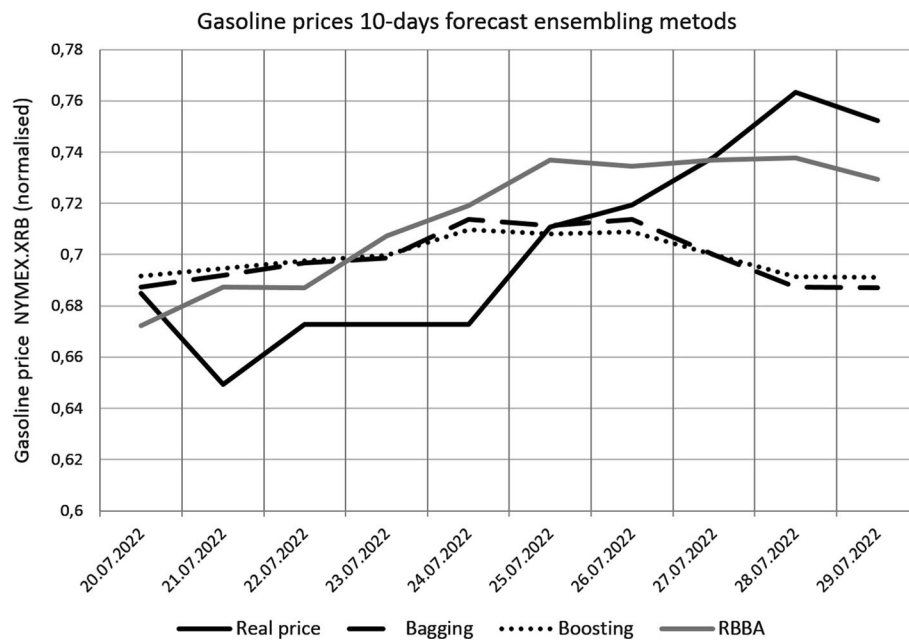$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}},\tag{17}$$

$$\Delta F(t)_{ens} = \pm\, t_\alpha\, \frac{RMSE_{F_{ens}}}{\sqrt{n}}.\tag{18}$$

where *RMSE* is the forecast error, i is the number of sample elements of the forecast residuals, n is the sample size of forecast residuals, x_i is the value of the time series of the forecast residuals, $\bar{x}$ denotes the arithmetic mean of x_i, $\Delta F(t)_{ens}$ is the forecast confidence interval, and t_α is the Student's t-distribution coefficient corresponding to the significance level α.

Quality evaluation of ensembles Below is the general result of the quality evaluation of ensembles for $\alpha = 0.05$, see Table 1 and Fig. 3.

Table 1 The final result of the quality evaluation of the forecast ensembling methods

Method	RMSE (the smaller the better)	Confidence interval (the smaller the better)	Forecast accuracy (the higher the better) (%)
Bayes optimal classifier	0.040325645	0.024994094	94.27
Bootstrap aggregating	0.042994277	0.026648129	93.89
Adaptive boosting	0.039093313	0.024230286	94.44
RBBA	0.026864154	0.016650575	96.18

Fig. 3 Gazoline prices forecast ensembling methods

Discussion

Evaluating the above, it is worth noting that the idea of the ideality of random residuals in itself is just a widespread hypothesis. In fact, when analyzing real data, we often observe periods of the distribution of residuals that are not only extremely close to normal (for example, the dollar exchange rate in the Russian Federation during the 2009–2014 period) but also extended periods with significant asymmetry and frequent outliers (for example, the dollar exchange rate in the Russian Federation in the 1994–2000 period), when the residuals do not correspond to statistical characteristics of normality by most criteria, which means that they imply trends that are unaccounted for and dependencies missed or deliberately ignored by forecasters. The optimization criteria chosen by the authors, despite the increase in the quality of predictive ensembles, do not guarantee that the approximation residuals correspond perfectly to the normal distribution but are only an alarming red flag indicating a significant intensity of processes and trends that are overlooked in the forecasts. The main advantages of the proposed method include higher prediction accuracy and no

need to perform a quality test on the residuals of the forecast, which leaves out a whole mandatory stage from the standard methodology of statistical research. Owing to the optimization of the quality of residuals, we obtain a multidimensional, optimal result as early as the forecasting stage. However, because the proposed method uses criteria applicable exclusively for large samples, it is not applicable for small ones, which is its real downside. Over the past 120 years, statistics has come a long way as a separate scientific field, but the period of transition to the third information age, when computing resources became extremely cheap and new statistical methods gained high demand again, was practically ignored by it.

Despite the allure of the idea of an approximating "self-learning smart black box," which served the development and widespread use of multilayer neural networks of deep learning and the latest forecasting methods, we could not qualitatively predict the crisis of 2008, the Arab Spring, or the energy crisis of 2020–2022.

Conclusion

The idea of a multidimensional residual-based optimization approach is not limited to this evaluation method. As a penalty function, one can use third- and fourth-order deviations from the normal moments, as well as higher-order estimates, or other criteria- and non-criteria-based methods. It is evident that both the method of turning points and the Shapiro–Wilk test described in this article imply the evaluation of large samples. Meanwhile, the RBBA method remains suitable for samples with a small number of elements, although in this case, it is necessary to use similar criteria designed for small samples. It is also worth noting that the RBBA method cannot be sensitive to any initial conditions or parameters for calculating the forecast because, unlike stochastic bootstrap methods, which depend on the initial time, it uses statistical criteria and gradient methods.

It is also worth noting that the quality criteria of the residuals chosen by the authors are far from a comprehensive assessment, as parameters such as correlations of deep orders, higher moments of distributions, and systematic external asymmetry are clearly not considered in the penalty function, which is the main criterion for RBBA optimization. Perhaps, to improve the quality of forecasts, it is necessary to add other measures that primarily characterize the presence of unaccounted trends, hidden dependencies, and multiplicative factors.

Declarations

Competing interests
The author declares that she has no competing interests.

References

Bienaymé IJ (1874) Sur une question de probabilités. Bull Soc Math France 2:153–154

Breiman L (1996) Bagging predictors. Mach Learn 24(2):123–140

Brockwell PJ, Davis RA (2002) Springer texts in statistics (STS). In: Introduction to time series and forecasting, 2nd edn. Springer, New York

Corana A (2015) Architectural evolution of nvidia gpus for high-performance computing. Technical Report 150212, IEIIT-CNR, Genova

Durbin J, Watson GS (1950) Testing for serial correlation in least squares regression: I. Biometrika 37(3/4):409–428

Emer E (2018) Boosting (AdaBoost algorithm). https://math.mit.edu/rothvoss/18.304.3PM/Presentations/1-Eric-Boosting304FinalRpdf.pdf. Accessed 10 Oct 2018

Fienberg SE, Lazar N (2001) William sealy gosset. In: Statisticians of the centuries. Springer, New York

Gashler M, Giraud-Carrier C, Martinez T (2008) Decision tree ensemble: Small heterogeneous is better than large homogeneous. In: 2008 Seventh international conference on machine learning and applications: 11–13 December 2008; Diego, CA, USA, pp 900–905

Goodfellow I, Bengio Y, Courville A (2016) Back-propagation and other differentiation algorithms. In: Deep feedforward networks. MIT Press, pp 200–220

Hashem S, Schmeiser B, Yih Y (1994) Optimal linear combinations of neural networks: an overview. In: Proceedings of 1994 IEEE international conference on neural networks (ICNN'94): 02 July 1994. Orlando, FL, USA, pp 599–614

Hastie T, Rosset S, Zhu J, Zou H (2009) Multi-class adaboost. Statistics and Its. Interface 2(3):349–360

Hilbert M (2020) Digital technology and social change: the digital transformation of society from a historical perspective. Dialog Clin Neurosci 22(2):189–194

Kazakevičius KA (1988) Approximate formulas for statistical processing of mechanical test results. Indu Lab Diagn Mater 54(12):82–85

Kendall MG (1973) Time series. Griffin, London

Kou G, Xu Y, Peng Y, Shen F, Chen Y, Chang K, Kou S (2021) Bankruptcy prediction for SMEs using transactional data and two-stage multiobjective feature selection. Decis Support Syst 140:113429

Kuncheva LI, Whitaker CJ (2003) Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. Mach Learn 51(2):181–207

Li T, Kou G, Peng Y, Philip SY (2022) An integrated cluster detection, optimization, and interpretation approach for financial data. IEEE Trans Cybern 52:13848–13861

Liu Y, Yao X (1999) Ensemble learning via negative correlation. Neural Netw 12(10):1399–1404

McCartin BJ (2006) *e*: The master of all. Math Intell 28(2):10–21

Mitchell T (1997) Machine learning. McGraw-Hill Science, New York

Rich A, Jeffrey D (1996) Function evaluation on branch cuts. ACM SIGSAM Bull 30(2):25–27

Salman R, Alzaatreh A, Sulieman H, Faisal S (2021) A bootstrap framework for aggregating within and between feature selection methods. Entropy (Basel) 23(2):200

Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples). Biometrika 52(3/4):591–611

Stigler SM (1981) Gauss and the invention of least squares. Ann Stat 9(3):465–474

Transtrum MK, Sethna JP (2012) Improvements to the Levenberg-Marquardt algorithm for nonlinear least-squares minimization. arxiv, pp 1–32

Verhulst P-F (1838) Notice sur la loi que la population poursuit dans son accroissement. Corresp Math Phys 10:113–121

Wen F, Xu L, Ouyang G, Kou G (2019) Retail investor attention and stock price crash risk: evidence from China. Int Rev Financ Anal 65:101376

Wetherill GB (1981) Intermediate statistical methods. Chapman and Hall, New York

Zhang W, Zhou Y (2020) Level-set functions and parametric functions. In: The feature-driven method for structural optimization, 1nd edn. Elsevier, pp 9–46

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.