Prediction of short-term financial time series as a problem of adaptive extrapolation of finite spectrum functions with phase space model criteria

Alexander M. Akhmetshin, Tatiana V. Seliverstova, Oleg S. Antropov

Dniepropetrovsk National University, Naykova st., 13, Dniepropetrovsk, 49050 Ukraine
E-mail: akhm@mail.dsu.dp.ua

Abstract

Theoretical fundamentals and information aspects of method of extrapolation of short-term time series, based on idea of prolongation of analytical functions out of limits of initial temporal measurements area are stated in this work. The nature of predicted data is considered by using special criteria for the length of taken datasets, based on recurrence plots method, which gives an opportunity to estimate the predictability of observed time series via the determinism of the system.

Results of the experimental functional test of a method on examples of real financial series are also given. The description of algorithm and data segment length selection criteria is provided as well.

1. Introduction

The interest to the problems of prediction of economic and financial time series considerably boosted with development of new information technologies. The most widespread modern approaches in this direction are based on usage of different modifications of neuronetwork prediction and auto regression models on the basis of the concept of space of states with usage of the generalized algorithms of a Kalman filter.

The prediction of economic and financial time series is a rather important problem in modern economy. The modern economic conditions are characterized by constant changes of legislative bases of economic activity. Besides there are few corporations with durable periods of existence, in which the wide experience of activity is accumulated. So we have the situation of practical absence of qualitative learning samples for synthesis of parametric model or for training neuron networks. Due to these and other reasons the effectiveness of some mentioned methods comes to be rather poor, especially for short-term and non-stationary series.

These circumstances lead to development and active usage of various mathematical methods of polynomial or spline-interpolation on the given segment of time series with extrapolation on one step out of limits of this segment. The possible problems of such approach are also well known, and here the real success is possible only in special cases. It can be concluded that there is a necessity of development of new approaches that have more reliable theoretical basis, one of which can be a method based on the idea of adaptive extrapolation of functions with a finite spectrum, with respect to the nature of data in formation of data segments.

This direction was intensively developed in the years of 70-s', first of all with application to the problem of super resolution in spectral analysis, but then the interest to this direction, as it is seems to us, was primarily lost. By virtue of the above-stated reasons serious results haven’t been achieved then.

Our purpose is the usage of information possibilities of this method of extrapolation to the problems of prediction of short-term non-stationary economic series, taking into account the predictability of such series basing on phase space approach, when the known methods of extrapolation appear to be poor effective.

2. Algorithm

The idea of method is based on next statements and assumptions. First, the measurements of economic time series are made absolutely precisely (the concept about «forecast» and «noise» components of observed series is discarded). Secondly, it is supposed, that observed on the limited time interval T time series $s_T(t)$ represent a known segment of the function s(t) which has finite spectrum $S(\omega)$, i.e.

$$S(\omega) = \int_{-\infty}^{\infty} s(t)\exp(-j\omega t)dt \,, \qquad (1)$$

$$S(\omega) = 0 \,, |\omega| > \omega_{\lim}$$

As is known, the function s(t) with Fourier transform of sort (1) represents the analytical function, that gives theoretical possibility of usage of idea of an analytic continuation (extrapolation) of the known function $s_T(t)$ for limits of an interval T, on which it is defined. Thus it is necessary, that the function, obtained as a result of extrapolation, must coincide with initial one on this interval.

The practical possibility of consideration s(t) as to the function with a finite spectrum is given by that generally accepted fact, that any economic system is mostly sluggish (i.e. has low frequency) as well as deterministic, this fact allows to consider s(t) from a position of expression (1).

However we aren't pretty sure what exact value ω_{lim} has. This circumstance forces to use steadier iterative algorithms of adaptive extrapolation.

The practical solution of the problem of extrapolation is founded on usage of the following properties of finite functions: the spectrum limited function has unlimited duration in time, the time bounded function has unlimited spectrum width.

The sequence of operations, which is necessary to fulfill before the usage of algorithm of adaptive extrapolation are next:

1. Selection of length of data segment used for prediction. For this purpose we propose phase space method criteria.

2. Usage of Savitzky-Golay filter to select a low frequency component from the signal.

3. Formation of a complex signal, real part of it is the low frequency component, obtained on the previous step, and imaginary one taken as derivative of the low frequency component.

4. Application of a method based on idea of adaptive extrapolation to the obtained complex signal.

The analytical scheme of algorithm of a method of adaptive extrapolation includes the following stages:

1. Calculation of direct Fourier transform from a signal $s_T(t)$, with its designation as $S_0(\omega)$. The function $S_0(\omega)$ has unlimited width of a spectrum.

2. Selection of subregion ω_{lim} from all possible values of frequency area ω. Construction of the function $Z_0(\omega)$, which is equal

$$Z_0(\omega) = \begin{cases} S_0(\omega), \text{ for } |\omega| > \omega_{гр} \\ 0, \text{ for } |\omega| \le \omega_{гр} \end{cases}$$

The function $Z_0(\omega)$ can be get by multiplying $S_0(\omega)$ on the "window" function. $Z_0(\omega)$ is the first approximation to the required extrapolated function.

3. Calculation of inverse Fourier transform from $Z_0(\omega)$, with designation of it as $s_1(t)$. The duration of this function will be unlimited in comparison with $s_T(t)$.

4. Construction of function $z_1(t)$, which coincides with the function $s_T(t)$ in limits of a preset measuring interval T, i.e.

$$z_1(t) = \begin{cases} s_T(t), \ t \in T \\ s_1(t), \ t \notin T \end{cases}.$$

5. Calculation of direct Fourier transform from $z_1(t)$, with designation of it as $S_1(\omega)$. The first iteration of algorithm came to an end.

6. Repeat items 2-5 until the given accuracy ε will be reached:

$$\int_0^T [s_T(t) - z_n(t)]^2 \, dt \le \varepsilon$$

The quantity of required iterations can reach several thousands. The possibility of obtaining of solution is predefined by efficiency of used in the item 2 corrected localizations of estimations of required parameters.

3. Data length selection criteria

It is obvious, that uncritical application of such method, especially to short-term economic data can lead to pitfalls. So let's consider the nature of given data and define a criteria for the appropriate choice of datasets and its lengths using nonlinear dynamics systems approach.

The proposed criteria is based on recurrence plots (RP) method and its quantification. Method was originally introduced to visualize the time dependent behavior of the dynamics of systems, which can be pictured as a trajectory $\vec{x}_i \in R^n$ (i=1,…,N) in the n-dimensional space. It represents the recurrence of the phase space trajectory to a certain state, which is a fundamental property of deterministic dynamical systems. The main step of visualization is the calculation of matrix

$$R_{ij} = \Theta(\varepsilon_i - \|\vec{x}_i - \vec{x}_j\|), \quad i, j = 1,..., N$$

where ε_i is a cutoff distance, $\|\cdot\|$ is a norm, $\Theta(x)$ - Heaviside function. The phase space vectors for one-dimensional time series u_i from observations can be reconstructed using the Taken's time delay method, $\vec{x}_i = (u_i, u_{i+\tau}, ..., u_{i+(m-1)\tau})$, dimension m can be estimated via false nearest neighbors method.

For the estimation of the predictability of the system basing on its observational data we'll use some measures, defined using structures in RP.

The distribution of the lengths l of the diagonal structures in the RP is $P^{\varepsilon}(l) = \{l_i; i = 1,..., N_l\}$, where N_l is the absolute number of diagonal lines. Processes with stochastic behavior cause none or very short diagonals, whereas deterministic processes cause longer diagonals and less single recurrence points. So the ratio of points forming diagonal structures to all recurrence points

$$DET = \frac{\sum_{l=l_{\min}}^{N} l P^{\varepsilon}(l)}{\sum_{i,j}^{N} R_{ij}^{m,\varepsilon}}$$

is introduced in [2,4] as a measure for the predictability (determinism) of systems; lines formed by the tangential motion of the phase space trajectory are excluded.

While testing the time series through varying the lengths of datasets, it could be seen, that the distribution of determinism rate for a rather short non-stationary time series was mostly unimodal. Some results are shown in the next table:

Length of data	DET for initial data	DET for filtered data	Corresponding resulting figure
28	0.702	0.831	2
33	0.880	0.942	
38	0.854	0.921	
62	0.766	0.937	6
72	0.865	0.960	
82	0.850	0.952	

As there is no any special recommendation for choice of ε and m , false nearest neighbors method was used to estimate systems' dimensions, and was taken such threshold, that DET gained its clear maximum in single point of the chosen rather short data interval for prediction.

The best prediction results were achieved when using datasets with such lengths, where determinism had gained its local maximum, in every case the varying dataset is taken before the point, where prediction is started (conditions of prediction, such as filter and frequencies, are considered the same).

4. Some results

Here are some results of operation of described method. The length of data segments were chosen in accordance with appropriate criteria.

For this purpose at first we'll consider an economic series describing metrics of the currency proceeds by one of publicity agents on metrics of the quarter reporting for 3,5 years. The corresponding economic series is shown in a fig. 1.

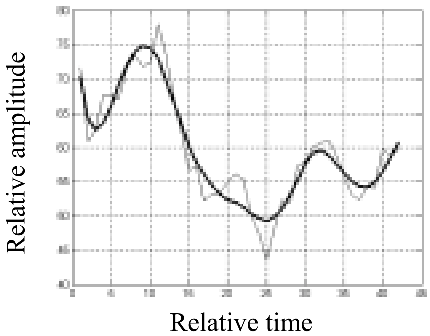

Fig. 1. Metrics of the currency proceeds of one of publicity agents quarter reporting for 3,5 years.

After application of a method of adaptive extrapolation we'll get results shown in fig. 2. The solid curve on the figure corresponds to the result of extrapolation, and dashed one shows the real behavior of the system.

As it can be seen from fig. 2, obtained as a result of extrapolation curve (real curve is shown as well), it goes downwards, and then will be lifted upwards.

That is the way obtained results mirror the common behavior of the function.

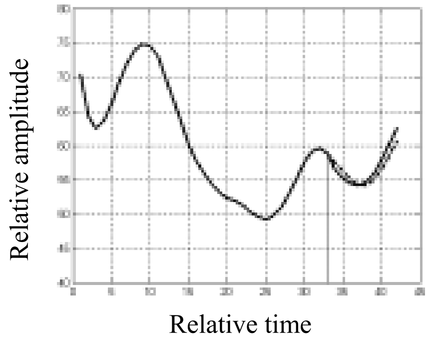

Fig. 2. Result of extrapolation.

Now we shall operate with economic series describing cost of one ton of coal in US dollars. The indicated economic series is shown in fig. 3.

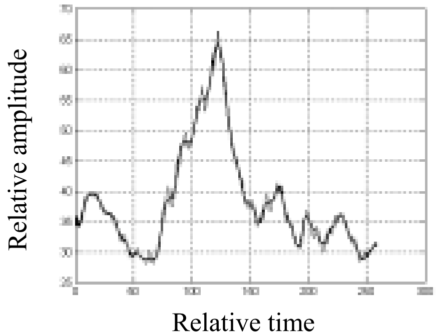

Fig. 3. Cost of one ton of coal in US dollars.

As we see this series is longer, than previous, and its behavior is highly non-stationary. The results of application of method of adaptive extrapolation are shown in fig. 4.

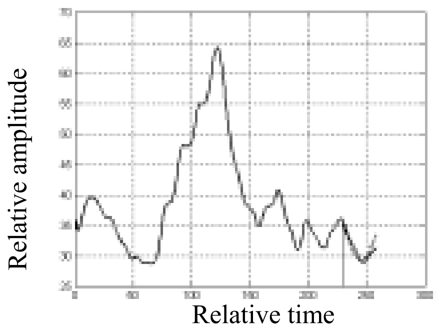

Fig. 4. Result of extrapolation.

As it can be seen from fig. 4, the curve, obtained as a result of extrapolation, is rather close to the real one. Thus for series of such length this method gives quite good enough results too, partly because of taking appropriate length of data segment, partly due to good fitting of analytical functions to the curve (frequencies are chosen rather precisely).

Let's consider an economic series describing dependence of cost of one ton of a grain in US dollars from productivity. The indicated economic series is shown in fig. 5.

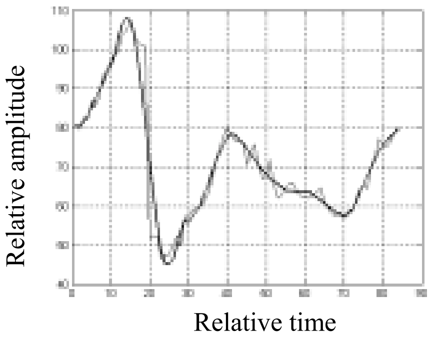

Fig. 5. Cost of one ton of a grain in US dollars depending on productivity.

This series is shorter than previous, but longer than first ones. After application of method of adaptive extrapolation we'll get results shown in fig. 6. The solid curve on the figure corresponds to the result of extrapolation, and dashed one shows the real behavior of the system.

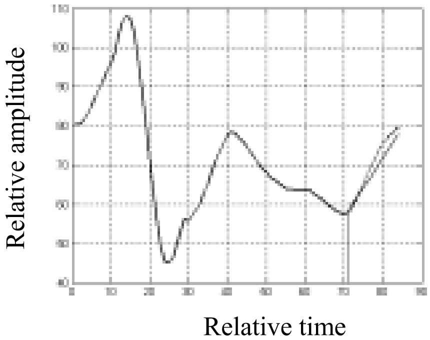

Fig. 6. Result of extrapolation.

Let's take one more short-term economic series, which is shown in fig. 7.

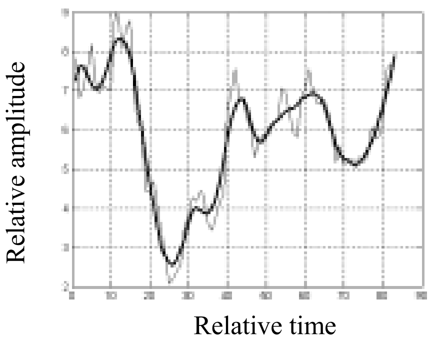

Fig. 7. Real economic series

After application of a method of adaptive extrapolation we'll get results shown in fig. 8.

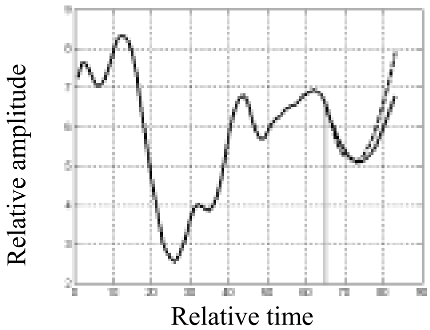

Fig. 8. Result of extrapolation.

For the third series the result of extrapolation is very close to the real function. For the fourth one we have received a correct direction. Thus, the results of extrapolations obtained for two last series mirror a general tendency of direction of assumed function rather precisely.

5. Conclusions

Thus, the method of adaptive extrapolation with appropriate phase space model criteria is a useful addition to widespread methods of extrapolation of values of time series. This method is mostly expedient for usage in the purposes of extrapolation of short-term series, when the efficiency of other well known methods becomes doubtful. The central problem of a method along with the choice of data segment is coupled to the choice of limits of spectrum area for the Fourier transforms. In each case the decision must be unique.

From the theoretical point of view, this method opens the possibility of usage of mathematical apparatus of the theory of analytical functions with the limited spectrum, in areas, where it yet hasn't been used. The described advantages of this method allow using it widely in practice for the problems of prediction of short-term economic and financial time series.

References

1. Papoulis A. A new algorithm in spectral analysis and band-limited extrapolation. IEEE Transactions on Circuits and Systems, 1975, №9, pp. 735-742.
2. Marwan N., Meinke A., Extended recurrence plot analysis and its application to ERP data, International Journal of Bifurcation and Chaos, 14(2), 2004
3. Timmer Y., Modeling volatility using state space models. – Int. J. Neural System. – 1997, v.8, № 4, pp. 678-684.
4. Webber Jr C.L., Zbilut J.P., Dynamical assessment of physiological systems and states using recurrence plot strategies. Journal of Applied Physiology 76(2), 1994, 965-973.
5. Zhang G., Patuwo B., Hu M.U. Forecasting with articial neural networks: The state of the art. – Int. J. of Forecasting, 1998, v.14, №1, pp. 35-62.