Scientific Inquiry, vol. 8, no. 2, 2007, pp. 183 – 194 *IIGSS Academic Publisher*

QUALITY IMPROVEMENT METHODS FOR STATISTICAL FORECASTS

DAINA ŠĶILTERE[a], JURIS KRASTS[b]

[a] *Department of economical system management theory and methods, University of Latvia, 5, Aspazijas bulv., Riga, LV-1050, Latvia;* [b] *Department of economical informatics, University of Latvia, 5, Aspazijas bulv., Riga, LV-1050,* Latvia

(Received May 26, 2006; In final form January 18, 2007)

The quality of statistical forecasts depends on the supplied data, among to other factors. The issues discussed in this work are quality problems of time series and statistical forecasts and possible solutions, to improve the quality. The main problems are the length of time series, stability and homogeneity of time series, random noise and decrease of value of the data in the course of time. The authors suggest iterative methods for excluding random noise or lessening its impact and time series discounting techniques.

Keywords: quality of forecasting data, time series, time series random noise, random noise reducing methods, time series discounting

DATA USED IN STATISTICAL FORECASTING

Increasing complexity of social life, rapid development of manufacturing, increasing dependency between economics and the politics and globalization have led to a more difficult forecasting process and necessity to develop forecasts with a higher reliability. The quality and speed of management decisions depends largely on the ability of managers to mine, analyze and interpret information.

The main requirements for developing justified complex forecasts are knowledge of forecasting theory, information about the study object and most influent factors on it. Increasing complexity of chosen forecasting methods and models requires larger amount and scope of data. It is important to remember, that forecasting results in data too.

Forecasting specialists often complain about lack of data, but, as stated in theoretical cybernetics, it is not possible to obtain all the data about all the aspects of the object. Nevertheless it is advised to achieve a compromise – use the most relevant information in widest scope possible.

The quality of forecasts largely depends on entireness, reliability and precision of data.

The data sets used in forecasting consist of:

- retrospective data about the study object,
- similar object development data (local and foreign experience);
- data about the forecasting background:
    - about related companies and institutions (partners, competitors),
    - about the economical, social and ecological situation locally and abroad,
    - about the law,
    - about the scientific achievements,

ISSN 1552-1222
http://www.iigss.net/Scientific-Inquiry/mission.html

- expert judgment about the past and possible future development of the study object.

When using statistical forecasting methods, the data sets usually consist of time series describing the study object and related factors. Following tasks can be solved by using time series analysis:

- determine the tendencies of the process – find the main tendency or trend and estimate random noise;
- determine and analyze periodical fluctuations (seasonality);
- study causal relationship between processes and occurrences, that show correlation between different time series;
- develop development model of the study object;
- forecast future development of the object, process or occurrence.

Qualitative forecast development largely depends on entireness, stability, reliability and precision of data. We will contribute to solving several time series quality problems. The main problems are the length of time series, stability and homogeneity of time series, random noise and decrease of value of the data in the course of time.

LENGTH OF TIME SERIES

To successfully analyze dynamic processes and forecast them the data used has to be entire, time series must be of a sufficient length according to defined forecasting tasks.

The definition of required time series length is not an unambiguous and easy to solve issue. Following concerns have to be taken into account to define the required length of time series:

♦ the forecasting task, for example, when analyzing and forecasting periodical fluctuations a time series of three full periods is advisable, that's why the required amount of information when forecasting seasonality is three years by month or quarters,

♦ forecasting skyline, for example, when developing long-term forecasts a longer time series is necessary in comparison with short-term forecasts,

♦ time series characteristics, for example, comparability of time series levels,

♦ requirements of the chosen research and forecasting methods, for example, when using statistical methods for analyzing seasonality a time series of at least four periods is required, but when developing regression models, the time series must be several times longer than the number of independent variables (T. A. Дуброва, 2003, p.13).

At the start of the forecasting process the required length of the time series can be only estimated, it will be defined precisely during the process of modeling and forecasting. There should be no qualitative changes in the time span that includes the time series. As can be concluded from the formula for calculating the confidence limits for statistical interval forecasts, the confidence limits of interval forecasts are inversely proportional to the length of time series. That means that by increasing the time series length ten times, the forecast error will decrease approximately three times. From this point of view it would be recommended, to use as long time lines, as possible. It should be pointed out, that this theory is developed for a truly random sample. There has been research done of how this can be applied in time series, because time series is not a truly random sample. But, from another point of view, a long time series contains incomparable values, because the study object changes during the time, just like methods of calculating indices. There has to be a compromise between these two factors when defining time series length. The decision of the time series length is influenced by the type of the forecasting model and forecasting skyline. For example, if a linear model is used for short-term forecasting, there is no reason to use a long time series, because almost all processes maintain a linear pattern only in short-term. When using scientifically grounded and approbated nonlinear models, like S-type curves, and developing mid-term and long-term forecasts, time series must be as long as possible. In real life, the available length of time series is limited, because of lack of comparable levels.

STABILITY OF STATISTICAL RATIOS

A significant demand when using statistical forecasting methods, like trend models and regressions, is the stability of the calculated ratios and conclusions.

One of the best interpretations of stability was made by professor O.Krastiņš: statistical ratios are stable, if they don't change a lot after acceptable changes have been made to the whole data set. The statistical ratio is dynamically stable (in time), if by calculating it on different time moments or periods for the same object, the resulted ratios are very similar or change in a pattern. Furthermore, the changes in the ratio can be explained by quantitative or qualitative changes of the study object (O.Krastiņš, 1987, p.3).

The statistical ratio should be evaluated as instable, if its change in different parts of time series (territories) is chaotic. These changes can not be described with territorial specifics or live changes.

Stability becomes important when evaluating different trend models (result of analytical smoothing) and causation models (regression). The models and their parameters should be dynamically stable to be used in analysis and forecasting.

There are two aspects to consider when measuring dynamical stability (И.И. Елисеева, М.М. Юзбашев, 2004, p.493). Firstly, stability as opposite to dispersion of time series and, secondly, stability of the direction of changes or stability of pattern is checked with correlation coefficient of the Spearman rank.

Main reasons that lessen the stability of statistical ratios:

- heterogeneous input data,
- random noise.

TIME SERIES HOMOGENEITY ANALYSIS

When using times series in forecasting methods, it is important to mention, that chronologically ordered data is not always a time series. The data must be comparable in a selected time period to be used in forecasting. That's why one of the prerequisites for data, to correctly represent real development in time, is homogeneity. Qualitative analysis should be made for non-comparable values.

A number of reasons exist for not-comparable values in chronologically arranged data:

- change of the economical or political system,
- change of the foreign exchange system,
- change of the territory representing the value, for example, moving the borders of states, regions, cities,
- classification changes,
- different observation moment or interval,
- new observation recording or calculating methodology,
- changed recording units,
- terminology changes, for example, time series describing the number of SME's can change significantly, if the definition of a SME is changed, a term has to be maintained during the whole accounting period,
- changed ratio unit of measure,
- changed value of absolute prices, that's why prices are calculated in comparable prices,
- structural changes, for example, merged or divided enterprises.

The problem of incomparable levels in time series can be solved in different ways depending on the forecasting task. In most cases it is done by recalculating or correcting time series. But such a solution doesn't provide the desired precision, decreases the value of data and burdens further analysis and forecasting.

Construction of a trend usually starts by isolating homogenous periods in the time series. The pattern in the last homogenous period is prior in forecasting.

Methods used for isolating homogenous periods

- qualitative or logical analysis,
- statistical methods.

Qualitative or logical analysis of the economic process is the first process done, to isolate homogenous periods in time series. Relevant changes can be triggered by new technology, founding a new enterprise, starting production of a new product, reorganization, law changes etc. Time series can be analyzed visually, too.

Homogenous periods can be also isolated by using statistical methods. Usage of a special criterion is recommended for such an analysis (Применение регрессионных моделей…, 1970, p.14).

If there are no prior conditions, time series is split in two parts, so, that $n_1 = n_2$ or $n_1 < n_2$. These to periods of time series are homogenous, if

$$\frac{Q-1}{\sigma(Q)} < 3 ,$$

where

$$Q = \frac{n_1(n_2-3)}{n_2(n_1-1)} \cdot \frac{\sigma_1^2}{\sigma_2^2} , \quad \sigma(Q) = \sqrt{\frac{2(n_1+n_2-4)}{(n_1-1)(n_2-5)}} ;$$

n_1 and n_2 – number of observation in the first and second part of the time series,

$\sigma_1^2 ; \sigma_2^2$ - dispersions: $\sigma_1^2 = \frac{\sum_{t=1}^{n_1}(y_t - \bar{y}_1)^2}{n_1}$ and $\sigma_2^2 = \frac{\sum_{t=n_1+1}^{n_2}(y_t - \bar{y}_2)^2}{n_2}$,

$\bar{y}_1 ; \bar{y}_2$ - average time series values in first and second part.

This criterion allows testing a hypothesis, if the differences in time series values are truly random, or these difference are so essential, that they can't be explained only with the influence of random factors. That's the case, when the inequality doesn't solve.

Fisher criterion is recommended to check homogeneity of time series (Применение регрессионных моделей…, 1970, 14. lpp.).

If it is necessary to combine incomparable values in a time series, special algorithms and software is used. As a result incomparable values are transformed by using recalculation coefficients, for example, price index and base year methodology.

TIME SERIES RANDOM NOISE AND ITS DETECTION

Time series used for modeling often include non-typical values called random noise. This random noise has negative impact on model parameters and reduces practical use of forecasts.

Random noises are all the values, which do not correspond with study object potential patterns and influence the result of statistical methods.

Sometimes random noises are called anomaly points, sharply different measures or artifacts (Latin. arte – artificial). We will use the term „random noise" or „uncharacteristic point", because these are the widely used terms in special literature.

There are numerous reasons that can cause random noise in time series – sampling errors, data mining and processing mistakes. The mistakes named are technical or of the first kind. These mistakes can be found and prevented. Random noise can arise from the influence of different objective factors which appear randomly. These are the errors of the second kind, which can't be avoided.

Reduction of random noise (by substitution or smoothing) is an essential part of preprocessing data, because such a data with random noise can deform the analysis and forecast results.

Random noise can also represent real process development, for example, rapid changes in exchange rates or share prices. Such fluctuations can be replaced with theoretical values, if statistical methods and models are used and real value differences against theoretical values are calculated.

Of course, there are cases, when time series values, which are very far from the theoretical trend, are not random noise, but part of the real time series pattern. Such a case is an indication of inadequately chosen trend. Sometimes random noise is caused by an unusual combination of different factors. In any case random noise detection methods should be combined with qualitative analysis of the study process.

There are different statistical methods for determining random noise, which where developed for analysis of empirical distributions, for example, Grubbs method (F.Grubbs, 1950, p. 27-58), Tietjen-Moore method (G.Tietjen, H.Moore, 1972, p. 583–597). All these methods, with exception of series criterion, are discussed in the work (С.А. Смоляк С.А., Б.П. Титаренко, 1980, p. 91-107). These methods are not intended for determination of random noise.

Irvin method is used to define random noise in time series (М.С.Красс, Б.П.Чупринов, 2004, p.405-407; Экономико-математические методы…, 2002, p.148).

Irvin method. Factual values of the Irvin criteria are calculated for the time series $y_1, y_2, ..., y_t, ..., y_n$:

$$\lambda_t = \frac{|y_t - y_{t-1}|}{\sigma_y},$$

where y_t – time series value,

t – period $t = 2, 3, ..., n$,

σ_y – standard deviation $\sigma_y = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \bar{y}_t)^2}{n-1}}; \bar{y} = \frac{\sum_{t=1}^{n} y_t}{n}$.

The calculated values $\lambda_2, \lambda_3, ..., \lambda_n$ must be compared to the critical values of the Irvin test λ_α and, if $\lambda_t > \lambda_\alpha$, then the according time series value y_t with a probability of $1-\alpha$ can be considered random noise.

The critical values of Irvin test at the significance level of $\alpha = 0,05$ or a error of 5 % are $\lambda_\alpha = 2,8$, if $n = 2$; $\lambda_\alpha = 2,3$, if $n = 3$; $\lambda_\alpha = 1,5$, if $n = 10$; $\lambda_\alpha = 1,3$, if $n = 20$ e.c.

Krastins method. One of the practically best applicable methods has been developed by prof. O.Krastiņš. The method was initially developed for variation series, but can be applied for time series, too (O.Krastiņš, 1987, p.6). The adopted method for time series is build on an assumption, that the distribution of deviations $\varepsilon_t = y_t - \tilde{y}_t$ corresponds to the normal distribution. The standard deviation must be calculated $S_{\tilde{y}}$.

The empirical t coefficient is calculated:

$$t = \frac{|\varepsilon_t - \bar{\varepsilon}_t|}{S_{\tilde{y}}},$$

where $\bar{\varepsilon}_t$ - average deviation ε_t.

If $t > t_\alpha$, where t_α - critical values of the Students distribution, the hypothesis is rejected. It means, that the deviation ε_t doesn't has a random nature and the time series observation y_t can be considered random noise.

It is easier to determine the smallest relevant distinction $t_\alpha \cdot S_{\tilde{y}}$. If any of the deviations ε_t exceeds $t_\alpha \cdot S_{\tilde{y}}$, then the corresponding time series observation y_t can be considered a random noise.

The quality of forecasts can be improved by excluding the values with random noise of lessening their impact:

- excluding separate time series values or whole parts,
- minimizing the impact of values with random noise:
  - random noise values are substituted with the average value of two adjacent values,
  - random noise values are substituted with a theoretical value of a trend model,
  - weight coefficients are applied to random noise values, these coefficients are calculated as a function from deviations.

EXCLUDING AND LESSENING THE IMPACT OF RANDOM NOISE

There are two alternate decisions that are recommended in literature to improve the results of forecasting:

- the questionable value or random noise is accepted as a logical value and is being used in calculation as a wholesome value,
- the questionable value or random noise is excluded from calculation as faulty.

It is done by using weight coefficients, which are assigned to the corresponding value – $\beta_t = 1$, if the value remains, $\beta_t = 0$, value is excluded from the time series. The disadvantages of this method are discussed in depth by prof. O.Krastiņš (O. Krastiņš, 1987, p.7-10).

The authors recommend an iterative procedure, to exclude or lessen the impact of random noise.

<u>The iterative random noise exclusion method</u>

1. Determination of the initial model parameters values $A_0 = \varphi(y_t, t)$, calculation of theoretical values $\tilde{y}_{t0} = f(A_0, y_t, t)$ and standard deviation $S_{\tilde{y}_0}$
2. Calculation of deviations $\varepsilon_{t_0} = y_t - \tilde{y}_{t_0}$
3. Assigning values to weight coefficients:

   $\beta_t = 0$, where $\left|\varepsilon_{t_0}\right| > t_\alpha S_{\tilde{y}_0}$

   $\beta_t = 1$, where $\left|\varepsilon_{t_0}\right| < t_\alpha S_{\tilde{y}_0}$
4. Assessment of model parameters $A = \varphi(y_t, t, \beta_t)$, calculation of theoretical values $\tilde{y}_t = f(A, y_t, t, \beta_t)$ and $S_{\tilde{y}}$
5. Calculation of deviations $\varepsilon_t = y_t - \tilde{y}_t$
6. Verification: if $m+1 < k$, then $\varepsilon_{t_0} = \varepsilon_t$, $S_{\tilde{y}_0} = S_{\tilde{y}}$ and do the next iteration (return to point 3.); if $m+1 = k$, then follow to the next step. Here m – number of excluded y_t, k – the allowed number of excludable values.
7. Results of the modeling $\tilde{y}_t = f(A, y_t, t, \beta_t)$.

An alternative approach to exclusion of random noise is lessening its impact. In this case the weight coefficient β_t is defined as a function from deviations ε_t. If the deviations ε_t are distributed according to the normal distribution, β_t can be found by using the differential function of the normal distribution. The impact lessening of random noise is done by using the iterative process.

Iterative random noise impact lessening method

1. Assessment of model $\tilde{y}_{t0} = f(A_0, y_t, t)$ parameters $A_0 = \varphi(y_t, t)$ un $S_{\tilde{y}_0}$
2. Calculation of deviations $\varepsilon_{t0} = | y_t - \tilde{y}_{t0} |$
3. Defining weight coefficients $\beta_t = e^{-\frac{\varepsilon_{tr}^2}{S_{\tilde{y}_0}^2}}$
4. Assessment of $A = \varphi(y_t, t, \beta_t)$ model $\tilde{y}_t = f(A, y_t, t, \beta_t)$ and calculate $S_{\tilde{y}}$
5. Calculation of deviations $\varepsilon_t = | y_t - \tilde{y}_t |$
6. Verification: $\left| \sum_{t=1}^{n} \varepsilon_t - \sum_{t=1}^{n} \varepsilon_{t_0} \right| < \Delta$, where Δ - precision of calculation. If the inequality doesn't proves true, initial values are replaced with values $\varepsilon_{t_0} = \varepsilon_t$, $S_{\tilde{y}_0} = S_{\tilde{y}}$ and next iteration is performed (return to point 3), if inequality proves true next step is executed.
7. Results of modeling $\tilde{y}_t = f(A, y_t, t, \beta_t)$

Weight coefficients can be determined not just by statistical methods, but also by using qualitative methods (expert survey). It allows combining different forecasting methods.

The results of the methods described can be seen in an experimental example (table 1). Long time series are used in this example. Such a time series with random noise are characteristic for non-aggregated ratios, for example, sales without seasonality. To smooth the time series a non-linear power trend model $\tilde{y}_t = at^b$ is used. Random noise exclusion weight coefficients are marked as β_t^*, but impact lessening coefficients as β_t^{**}. By analyzing the results a conclusion can be maid, that iterative method of impact lessening outputs better results from the point of view of forecasting. The long forecasting skyline has been chosen to visualize the difference in results.

Table 1: Forecasting with random noise exclusion and random noise impact lessening iterative methods

t	y_t	$\tilde{y}_t$	$\tilde{y}_t - S_{\tilde{y}}$	$\tilde{y}_t + S_{\tilde{y}}$	β_t^*	$\tilde{y}_t^*$	β_t^{**}	$\tilde{y}_t^{**}$
1.0	2.60	1.93	-0.504	4.36	1	1.915	0.9259	1.975
2.0	2.60	2.59	0.158	5.02	1	2.590	0.9999	2.619
3.0	3.50	3.08	0.647	5.51	1	3.091	0.9701	3.089
4.0	4.40	3.48	1.048	5.91	1	3.504	0.8659	3.472
5.0	1.30	3.83	1.396	6.26	0	3.862	0.3395	3.802
6.0	3.00	4.13	1.705	6.56	1	4.181	0.8040	4.095
7.0	0.11	4.42	1.986	6.85	0	4.471	0.0432	4.360
8.0	3.80	4.67	2.244	7.10	1	4.739	0.8785	4.604
9.0	11.40	4.91	2.485	7.34	0	4.989	0.0008	4.830
10.0	4.90	5.14	2.711	7.57	1	5.223	0.9902	5.042
11.0	5.10	5.35	2.924	7.78	1	5.445	0.9891	5.241
12.0	5.90	5.56	3.127	7.99	1	5.655	0.9802	5.430

13.0	5.50	5.75	3.320	8.18	1	5.856	0.9895	5.610
14.0	5.70	5.93	3.504	8.36	1	6.048	0.9907	5.782
15.0	6.80	6.11	3.681	8.54	1	6.233	0.9227	5.946
16.0	10.70	6.28	3.852	8.71	0	6.411	0.0366	6.105
17.0	6.90	6.45	4.016	8.88	1	6.582	0.9656	6.257
18.0	7.40	6.61	4.175	9.03	1	6.748	0.8984	6.404
19.0	4.10	6.76	4.329	9.19	0	6.909	0.3019	6.547
20.0	4.17	6.91	4.479	9.34	0	7.065	0.2807	6.685
21.0		7.05				7.217		6.819
22.0		7.20				7.365		6.950
23.0		7.33				7.509		7.076
24.0		7.47				7.650		7.200
25.0		7.60				7.787		7.321

where $\tilde{y}_t = 1.926 \cdot t^{0.426}$, $\tilde{y}_t^* = 1.915 \cdot t^{0.436}$, and $\tilde{y}_t^{**} = 1.975 \cdot t^{0.407}$. Results of the forecasting by using random noise exclusion and impact lessening iterative methods are visualized in figure 1.

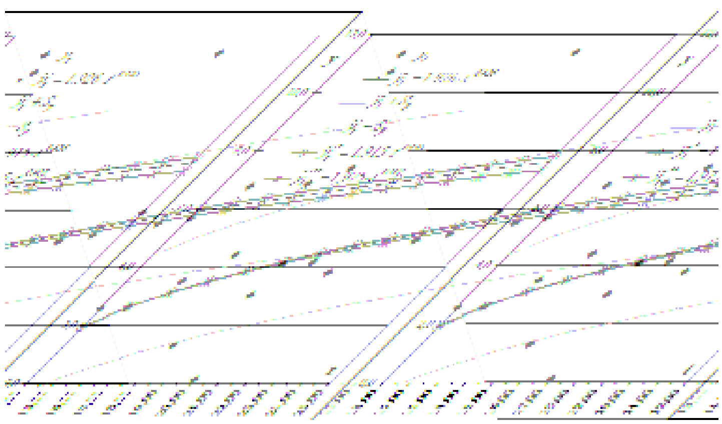

Figure 1. Smoothed time series and forecasts calculated with random noise exclusion and impactlessening iterative methods

TIME SERIES DISCOUNTING

When forecasting socially economical processes, the "value" of observations becomes important. Time series discounting can be used in connection with this issue.

Least squares method is based on a hypothesis, that the value of data is independent from time. That's why all the values in time series are used in calculations equally. But the fast scientific progress, new management strategies and other external factors have changed the patterns in time series. In such circumstances it is not justly to use long time series, because they include data about the development phase, which is irrelevant in trend modeling. The forecasting error increases by using shorter time series. A compromise is to use weight coefficients. A weight coefficient β_t is assigned to each time series value.

The following technique is usually used to determine discounting weight coefficients:

$\beta_t = \frac{t}{n}$; where $t = \overline{1,n}$. It is difficult to achieve desired results when using this technique, because similar weight coefficients are assigned to all time series values. For example, if time series consists of 10 observations, then the weight coefficients will be 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1. As it can be seen, last three or four time series weight coefficients vary a lot, even by a third. It is not acceptable in forecasting, because the several last periods are the most influential in forecasts and it is not correct to emphasize only the one last period.

Authors propose to calculate the weight coefficients for time series discounting by using following functions:

- half-logarithmic function $\beta_t = a + b\ln t$,
- logistic function $\beta_t = \frac{1}{c + ab^t}$,
- Pearl-Reed function $\beta_t = \frac{c}{1 + ae^{-bt}}$,
- integral function of the normal distribution $\beta_t = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{(t-\bar{t})^2}{2\sigma^2}} dt$.

Values of weight coefficients for time series discounting are listed in table 2. In most cases it is necessary to take three or four values in account. All the mentioned functions comply with this demand, but Pearl-Reed function allows to assign weight coefficients to almost half of the time series (figure 2).

Table 2: Weight coefficient values for time series discounting

t	$\beta_t = \frac{t}{n}$	Logistical curve	Pearl-Reed curve	Half-logarithmic curve	Integral function of the normal distribution
1	0.050	0.001	0.0001	0.0000	0.023
2	0.100	0.002	0.0002	0.2314	0.040
3	0.150	0.004	0.0007	0.3667	0.067
4	0.200	0.008	0.0020	0.4628	0.106
5	0.250	0.020	0.0058	0.5372	0.159
6	0.300	0.045	0.0169	0.5981	0.227
7	0.350	0.098	0.0480	0.6496	0.309
8	0.400	0.203	0.1289	0.6941	0.401
9	0.450	0.373	0.3026	0.7335	0.500
10	0.500	0.581	0.5599	0.7686	0.599
11	0.550	0.764	0.7886	0.8004	0.691
12	0.600	0.883	0.9162	0.8295	0.773
13	0.650	0.947	0.9698	0.8562	0.841
14	0.700	0.976	0.9895	0.8809	0.894
15	0.750	0.990	0.9964	0.9040	0.933
16	0.800	0.996	0.9988	0.9255	0.960
17	0.850	0.998	0.9996	0.9457	0.977
18	0.900	0.999	0.9999	0.9648	0.988
19	0.950	1.000	1.0000	0.9829	0.994
20	1.000	1.000	1.0000	1.0000	0.997

where the logistical curve is defined as $\beta_t = \frac{t}{n}$, the Pearl-Reed curve $\beta_t = \frac{1}{(c + a \cdot b^t)}$, the half-logarithmic curve is defined by $\beta_t = a + b \cdot \ln(t)$, and the integral function of the normal distribution $\beta_t = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$.

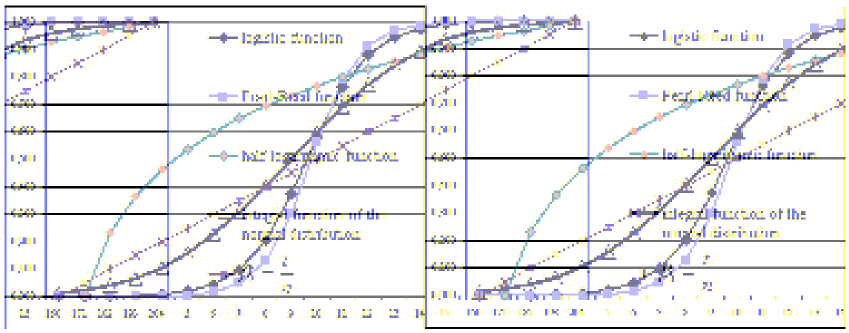

Figure 2. Time series discounting weight coefficients calculation functions

To illustrate the proposed discounting techniques and results an experimental example based on a long time series and long forecasting skyline was used. (Table 3 and figure 3). Forecasts were calculated by using the non-linear power trend model.

Table legends are as follows:

- $\tilde{y}_t$ - smoothed time series and forecast (without discounting);
- $\tilde{y}_t^{(1)}$ - smoothed time series and forecast, discounting made with weight coefficients calculated by logistical function;
- $\tilde{y}_t^{(2)}$ - smoothed time series and forecast, discounting made with weight coefficients calculated by Pearl-Reed function;
- $\tilde{y}_t^{(3)}$ - smoothed time series and forecast, discounting made with weight coefficients calculated by half-logarithmic function;
- $\tilde{y}_t^{(4)}$ - smoothed time series and forecast, discounting made with weight coefficients calculated by integral function of the normal distribution.

We can conclude (figure 3), that the best results can be achieved by using Pearl-Reed model.

Table 3: Forecasting results by using discounted time series

t	y_t	$\tilde{y}_t$	$\tilde{y}_t^{(1)}$	$\tilde{y}_t^{(2)}$	$\tilde{y}_t^{(3)}$	$\tilde{y}_t^{(4)}$
1.0	0.93	0.666	0.842	1.207	2.0211	1.7252
2.0	0.15	1.123	1.336	1.751	2.5788	2.2879
3.0	0.50	1.524	1.751	2.177	2.9739	2.6987
4.0	1.21	1.893	2.121	2.540	3.2905	3.0342
5.0	1.30	2.240	2.460	2.864	3.5590	3.3229
6.0	3.00	2.570	2.778	3.158	3.7946	3.5790

7.0	3.11	2.887	3.079	3.431	4.0059	3.8109
8.0	3.80	3.193	3.365	3.686	4.1984	4.0239
9.0	2.10	3.489	3.640	3.926	4.3759	4.2216
10.0	4.90	3.778	3.904	4.155	4.5410	4.4067
11.0	5.10	4.059	4.160	4.373	4.6958	4.5811
12.0	5.54	4.334	4.408	4.582	4.8416	4.7464
13.0	5.63	4.604	4.650	4.783	4.9798	4.9036
14.0	5.45	4.869	4.885	4.977	5.1113	5.0539
15.0	5.10	5.129	5.115	5.165	5.2368	5.1979
16.0	5.23	5.384	5.340	5.347	5.3569	5.3363
17.0	5.24	5.636	5.560	5.524	5.4723	5.4697
18.0	5.23	5.884	5.775	5.696	5.5834	5.5986
19.0	5.67	6.129	5.987	5.864	5.6906	5.7232
20.0	5.69	6.371	6.195	6.027	5.7941	5.8440
21.0		6.609	6.400	6.187	5.8944	5.9613
22.0		6.845	6.601	6.344	5.9915	6.0753
23.0		7.079	6.800	6.497	6.0859	6.1863
24.0		7.309	6.995	6.647	6.1777	6.2945
25.0		7.538	7.188	6.794	6.2670	6.4000

The conformity of the data with all the mentioned requests (time series length, comparability and homogeneity of time series values, lessened influence of time series, conformity of the data value) is validated during the first phase of time series analysis. Further analysis and forecasting can be performed only after this phase.

The problems, mentioned in this work, are not the only ones important in statistical forecasting. Forecast quality is highly dependent on objectivity, credibility and precision of the data. Nowadays the importance of data quality control has increased. It is linked to following circumstances. Firstly, the economy has become a more complex system, where wrong decisions can result in long-term negative consequences. Secondly, the fast development in use of information technologies for econometrical methods and integrated data processing systems has increased the requests for data quality. Errors in primary data can aggregate and influence the results of analysis. Thirdly, the amount of information has increased and therefore data credibility control has become very labor-consuming.

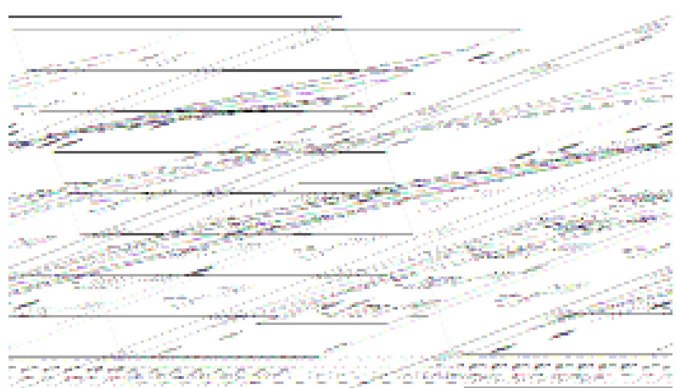

Figure 3. Smoothed time series and forecasts by using different discounting techniques

References

Krastiņš O. Ievads stabilo vērtējumu metodēs. - Riga: LVU, 1987 – 24 p.

Vasermanis E., Šķiltere D., Krasts J.. Prognozēšanas metodes. - Riga: LU, 2002 – 129 p.

Дуброва Т.А. Статистические методы прогнозирования.- Moscow: ЮНИТИ, 2003 – 206 p.

Елисеева И.И., Юзбашев М.М. Общая теория статистики. Moscow: Финансы и статистика, 2004 – 656 p.

Красс М.С., Чупринов Б.П. Математика для экономистов. - СПб.: St. Petersburg, 2004 – 464 p.

Крастинь О.П. Об исключении резко выделяющихся наблюдений и нормальном взвешиванию. // Математические методы в экономике. Вып.20. Riga, 1984 – p. 30-51.

Применение регрессионных моделей для прогнозирования экономических показателей. Под ред. А.С.Довбы. Moscow,1970 – 62 p.

Смоляк С.А., Титаренко Б.П. Устойчивые методы оценивания. Moscow: Статистика, 1980 – 208 p.

Экономико-математические методы и прикладные модели. Moscow: ЮНИТИ, 2002 – 391 p.

Grubbs F. Sample criteria for testing outlying observations. // Annals of Mathematical Statistics. 1950 vol. 21, p. 27-58.

Tietjen G., Moore H. Some Grubb's type statistics for the detection of several outliers. // Technjmetrics, 1972, vol. 14, p. 583 – 597.