

Improved Holt Method for Irregular Time Series

T. Hanzák

Charles University, Faculty of Mathematics and Physics, Prague, Czech Republic.

Abstract. The paper suggests an improvement of Holt method for irregular time series as was presented by Wright. The modification deals with problem of time-close observations, i.e. two subsequent observations with time distance much shorter than in average. If this situation occurs in time series when using the original Wright's formulae, one can obtain seriously wrong results. Simulation study is provided to compare the performance of the original and improved method.

1. Introduction

At the end of 50's, Holt and Winters proposed an ad hoc method for smoothing and forecasting time series with locally linear trend [Winters, 1960] nowadays referred usually as *Holt method*. They employed the exponentially weighted moving average approach to estimate the level and newly also the slope of the time series. The extension of this method for the case of seasonal time series (*Holt-Winters method*) was provided at the same time.

Holt method has achieved a broad popularity among forecasters due to its simplicity and good performance. Several modifications of this original method have appeared in literature and are used in practice. The version with damped linear trend which is said to give better results in longer forecasting horizons is well known [Gardner and McKenzie, 1985]. Wright [1986] has suggested a straightforward generalization of Holt method for the case of time series observed at irregular time intervals. He has extended the original smoothing formulae to this situation and has tested his method on several real time series. The formulae of classical Holt method and its extension for irregular time series by Wright are reminded in Section 2.

Concerning time series with missing observations, the time distance between two consequent observations is bounded from below by the concerned time unit. This bound is not dramatically lower than the average time spacing in the series. In contrast to this, if a general time irregularity is allowed, we can come to a situation which we call as a problem of *time-close observations*: two consequent observations have much shorter time distance, compared to the average time spacing in the series.

Due to the measurement noise or due to any other special character of data, these observations can have their values significantly different even if their times are almost the same. Thus a one-step slope between these two observations can be theoretically arbitrarily large. Using the Wright's extension of Holt method, this phenomenon will then appear more or less also in the time series of corresponding smoothed levels and can seriously affect the slope estimate and consequently cause a bias in forecasts for a quite long time period. Moreover, the slope estimate bias can occur even if the values of time-close observations are close to each other. Detailed discussion of this impact is provided in Section 3.

In Section 4 we suggest a reasonable modification of Wright's formulae to overcome these difficulties. This modification is not less intuitive or easy to understand than the original formulae. In author's opinion it has even a better interpretation. Moreover, as far as the resulting formulae are concerned, the modification means just adding one term in an updating formula for slope smoothing coefficient.

In Section 5 a simulation study is provided to check the impact of the suggested modification in practice. Several settings of time series generating scheme are employed to find out how the improvement depends on the configuration of time-close observations in data. Section 6 brings the conclusions of the paper.

2. Holt method for irregular data

Let us suppose that $\{y_t, t \in \mathbb{Z}\}$ is a (regular) time series with locally linear trend. The classical Holt method concerns its level S_t and slope T_t at time t . This means that the forecast $\hat{y}_{t+\tau}(t)$ of the future unknown observation $y_{t+\tau}$, $\tau > 0$, from time t is of the form

$$\hat{y}_{t+\tau}(t) = S_t + \tau \cdot T_t. \quad (2.1)$$

After a new observation y_{t+1} becomes available, the level and slope are updated using recurrent formulae

$$S_{t+1} = (1 - \alpha) \cdot (S_t + T_t) + \alpha \cdot y_{t+1}, \quad (2.2)$$

$$T_{t+1} = (1 - \gamma) \cdot T_t + \gamma \cdot (S_{t+1} - S_t), \quad (2.3)$$

where $\alpha \in (0, 1)$ is a smoothing constant for level and $\gamma \in (0, 1)$ a smoothing constant for slope. Formula (2.3) is a recurrent version of the exponential weighting

$$T_t = \gamma \cdot \sum_{j=0}^{\infty} (1 - \gamma)^j \cdot \tilde{T}_{t-j}, \quad (2.4)$$

where $\tilde{T}_k = S_k - S_{k-1}$ is the one-step slope from time $k - 1$ to k .

Wright [1986] suggested an extension of this method for the case of time series observed at irregular time intervals. Let $\{y_{t_n}, n \in \mathbb{Z}\}$ be such a time series, $t_{n+1} > t_n$, $n \in \mathbb{Z}$. *Wright* followed the idea of exponential weighting and generalized formulae (2.2) and (2.3) into

$$S_{t_{n+1}} = (1 - \alpha_{t_{n+1}}) \cdot [S_{t_n} + (t_{n+1} - t_n) \cdot T_{t_n}] + \alpha_{t_{n+1}} \cdot y_{t_{n+1}}, \quad (2.5)$$

$$T_{t_{n+1}} = (1 - \gamma_{t_{n+1}}) \cdot T_{t_n} + \gamma_{t_{n+1}} \cdot \frac{S_{t_{n+1}} - S_{t_n}}{t_{n+1} - t_n}, \quad (2.6)$$

where variable smoothing coefficients α_{t_n} and γ_{t_n} are updated in a recurrent way

$$\alpha_{t_{n+1}} = \frac{\alpha_{t_n}}{\alpha_{t_n} + (1 - \alpha)^{t_{n+1} - t_n}} \quad \text{and} \quad \gamma_{t_{n+1}} = \frac{\gamma_{t_n}}{\gamma_{t_n} + (1 - \gamma)^{t_{n+1} - t_n}}. \quad (2.7)$$

Again, formula (2.6) is a recurrent version of the underlying exponential weighting

$$T_{t_n} = \gamma_{t_n} \cdot \sum_{j=0}^{\infty} (1 - \gamma)^{t_n - t_{n-j}} \cdot \tilde{T}_{t_{n-j}}, \quad (2.8)$$

where

$$\gamma_{t_n} = \left[\sum_{j=0}^{\infty} (1 - \gamma)^{t_n - t_{n-j}} \right]^{-1} \quad (2.9)$$

plays the role of a normalizing factor and

$$\tilde{T}_{t_k} = \frac{S_{t_k} - S_{t_{k-1}}}{t_k - t_{k-1}} \quad (2.10)$$

is again the one-step slope from time t_{k-1} to t_k . Initial values of smoothing coefficients α_{t_n} and γ_{t_n} are taken as fixed points of formulae (2.7) with $t_{n+1} - t_n = q$, where $q > 0$ is the average time spacing of time series y :

$$\alpha_{t_0} = 1 - (1 - \alpha)^q \quad \text{and} \quad \gamma_{t_0} = 1 - (1 - \gamma)^q. \quad (2.11)$$

Initial values of S and T are constructed in a usual way just as in the regular case.

3. Problem of time-close observations

As was already said in Section 1, time step $t_{n+1} - t_n$ can be sometimes much shorter than the average time spacing q . But even if $t_{n+1} - t_n$ is approaching zero, the difference $y_{t_{n+1}} - y_{t_n}$ can be significant. One general reason consists in the measurement noise which causes that even if we make two observations in the exactly same time, we won't necessarily get the same observation values.

Moreover, some economic time series exhibit substantially large jumps in their true values during very short time periods. Prices of goods are a typical example of this type of data. *Wright* [1986] used several examples of irregular time series, one of them was the time series of world record times in one mile run. This is another example of a time series with time-close observations.

As will be seen from the next paragraph, it is even not necessary for the difference $y_{t_{n+1}} - y_{t_n}$ to be significantly nonzero and the short time distance $t_{n+1} - t_n$ itself can cause problems as well.

Let us look in a deeper detail how the presence of time-close observations affects results obtained by *Wright's* version of Holt method. From formulae (2.5) and (2.6) one can easily derive the following *error-correction* form of (2.6):

$$T_{t_{n+1}} = T_{t_n} + \frac{\gamma_{t_{n+1}} \alpha_{t_{n+1}}}{t_{n+1} - t_n} \cdot e_{t_{n+1}}, \quad (3.12)$$

where

$$e_{t_{n+1}} = y_{t_{n+1}} - \hat{y}_{t_{n+1}}(t_n) = y_{t_{n+1}} - [S_{t_n} + (t_{n+1} - t_n) \cdot T_{t_n}] \quad (3.13)$$

is the forecasting error from time t_n to time t_{n+1} . For $t_{n+1} \rightarrow t_{n+}$ (n and t_n are fixed), it is

$$\alpha_{t_{n+1}} \rightarrow \frac{\alpha_{t_n}}{\alpha_{t_n} + 1} > 0 \quad \text{and} \quad \gamma_{t_{n+1}} \rightarrow \frac{\gamma_{t_n}}{\gamma_{t_n} + 1} > 0, \quad (3.14)$$

see (2.7), and so

$$\frac{\gamma_{t_{n+1}} \alpha_{t_{n+1}}}{t_{n+1} - t_n} \rightarrow \infty. \quad (3.15)$$

Together with the fact that the forecasting error $e_{t_{n+1}}$ in (3.12) is not in any sense restricted to tend to 0 when $t_{n+1} \rightarrow t_{n+}$, it is possible that the difference between the original slope estimate T_{t_n} and the new one $T_{t_{n+1}}$ can be arbitrarily large. It is obvious that such a sudden bias in slope estimate will negatively influence the forecasts for a quite long time period ahead.

The intensity of the effect naturally depends on several factors:

- How close is the time step $t_{n+1} - t_n$ to zero.
- What are the values of α_{t_n} and γ_{t_n} . This is determined by the values of α and γ and the time structure of the series y from t_n back to its history.
- What is the value of the forecasting error $e_{t_{n+1}}$.

When $y_{t_{n+1}} - y_{t_n}$ is large compared to $t_{n+1} - t_n$ then *usually* the forecasting error $e_{t_{n+1}}$ is large as well. But as we have seen above, the true driver of the effect is the forecasting error $e_{t_{n+1}}$, not the difference $y_{t_{n+1}} - y_{t_n}$.

When using Holt method in practice, one usually chooses the values of smoothing constants α and γ as those minimizing certain criterion like *mean square error* (MSE). When the problem of time-close observations is present in a treated time series, searching for the optimal combination of α and γ can be reduced to searching for such a combination that eliminates the impact of time-close observations. It will usually happen that very small value of γ together with slightly higher value of α are optimal. But these values are probably not optimal for the rest of the series so a significantly worse overall predictive performance of the method can be expected. See Section 5 for more details.

Let us notice that the other methods for treating irregular time series with locally linear trend are not sensitive to time-close observations in data as the Wright's extension of Holt method is. We mean the double exponential smoothing for irregular data, see *Cipra* [2006], and the method based on discounted least squares estimation of linear trend, see *Cipra and Hanzák* [2008]. But these methods are less flexible since they use only one smoothing constant instead of two as the Holt method does. So it is worth to look for a version of Holt method which would be applicable to irregular time series without sensitivity to time-close observations.

4. Suggested solution

In previous section we have explained how exactly does the presence of time-close observations affect the results obtained by Wright's method. In this section we will give a solution to this problem.

The first possibility is to modify somehow the treated time series before we apply the forecasting method to it. One should go through the series and find all pairs of time-close observations in it. For such a pair, a reasonable solution is to transfer it to one single observation with its time and value taken as the arithmetic average of the individual times and values (geometrically, the segment is replaced by its middle). Such a preliminary data modification can be made in an automatic way.

If there is a measurement noise in our time series, the solution suggested above has one serious disadvantage. The measurement noise variance of the created joint observation is approximately just a half of its usual value. So correctly done, a double weight should be given to this observation in exponentially weighting formulae.

In the rest of this section we will suggest a modification of Wright's version of Holt method which will make it robust to time-close observations in the series. From the technical point of view, the problem is in formula (3.12) where the forecasting error $e_{t_{n+1}}$ is multiplied by the potentially infinite expression $\frac{\gamma_{t_{n+1}} \alpha_{t_{n+1}}}{t_{n+1} - t_n}$, see (3.15). But the problem can be seen also from the point of view of the weighting formula (2.8). Individual one-step slopes \tilde{T}_{t_k} are weighted here only according to the position of t_k on the time

axis. But one can expect that the reliability of \tilde{T}_{t_k} depends also on the length of the underlying time step $t_k - t_{k-1}$. As this time distance is approaching 0, the value of \tilde{T}_{t_k} is determined more by random measurement noise than by the true change in the time series level. So the weight given to \tilde{T}_{t_k} in (2.8) should be decreasing with decreasing value of $t_k - t_{k-1}$.

We suggest this weight to be equal to

$$\frac{t_k - t_{k-1}}{t_n - t_{n-1}} (1 - \gamma)^{t_n - t_k} \quad (4.16)$$

(the weight for $k = n$ is still equal to 1) so the formula (2.8) will be replaced by

$$T_{t_n} = \gamma_{t_n} \cdot \sum_{j=0}^{\infty} \frac{t_{n-j} - t_{n-j-1}}{t_n - t_{n-1}} (1 - \gamma)^{t_n - t_{n-j}} \cdot \tilde{T}_{t_{n-j}}, \quad (4.17)$$

where now, in contrast to (2.9),

$$\gamma_{t_n} = \left[\sum_{j=0}^{\infty} \frac{t_{n-j} - t_{n-j-1}}{t_n - t_{n-1}} (1 - \gamma)^{t_n - t_{n-j}} \right]^{-1}. \quad (4.18)$$

It is obvious that the recurrent formula (2.6) for the slope T remains unchanged. Just a different smoothing coefficient γ_{t_n} is used here, see (4.18). A recurrent formula analogous to (2.7) can be derived as well:

$$\gamma_{t_{n+1}} = \frac{\gamma_{t_n}}{\gamma_{t_n} + \frac{t_n - t_{n-1}}{t_{n+1} - t_n} (1 - \gamma)^{t_{n+1} - t_n}}. \quad (4.19)$$

So the whole suggested modification of the method relies just on adding the term $\frac{t_n - t_{n-1}}{t_{n+1} - t_n}$ into the formula (2.7) to change it into (4.19). The initial value for γ_{t_n} , following the Wright's concept of a fixed point, is exactly the same as given in (2.11). None of other parts of a practical implementation of the method is affected. Looking at formula (4.19), we see that for regular time series, our modified method turns into classical Holt method as well as the Wright's original version does.

Let us have a look at formula (3.12) again. Now, for $t_{n+1} \rightarrow t_{n+}$, it is

$$\frac{\gamma_{t_{n+1}} \alpha_{t_{n+1}}}{t_{n+1} - t_n} \rightarrow \frac{\gamma_{t_n} \alpha_{t_n}}{(t_n - t_{n-1})(\alpha_{t_n} + 1)} < \infty \quad (4.20)$$

so there is no more problem of seriously biased $T_{t_{n+1}}$. The weights suggested in (4.16) have also the following justification. Let us take $\gamma = 0$ and only finite versions of our smoothing formulae ($j = 0, \dots, J$) for a moment. Then formula (4.17) (using also appropriate finite version of (4.18)) will turn into

$$T_{t_n} = \frac{S_{t_n} - S_{t_{n-J-1}}}{t_n - t_{n-J-1}} \quad (4.21)$$

which seems to be more reasonable than the analogous finite version of (2.8):

$$T_{t_n} = \frac{1}{J+1} \cdot \sum_{j=0}^J \frac{S_{t_{n-j}} - S_{t_{n-j-1}}}{t_{n-j} - t_{n-j-1}}. \quad (4.22)$$

It is possible to modify all other methods familiar with Wright's version of Holt method exactly in the same way. It relates to Holt method with damped linear trend or exponential trend for irregular time series [Cipra, 2006] and Holt-Winters method for time series with missing observations [Cipra, 1995].

Seasonality in Holt-Winters method, either additive or multiplicative, has no impact on slope updating and so the suggested modification needs no special comments. With damped linear or exponential trend we can use exactly the same modified formula (4.19) as well. But now we are losing the justification provided by (4.21) which works exactly in the case of classical linear trend only.

In the case of damped linear trend one can think about replacing the formula (4.19) by

$$\gamma_{t_{n+1}} = \frac{\gamma_{t_n}}{\gamma_{t_n} + \frac{g(t_n - t_{n-1})}{g(t_{n+1} - t_n)} (1 - \gamma)^{t_{n+1} - t_n}}. \quad (4.23)$$

where $g(x) = \varphi \frac{1 - \varphi^x}{1 - \varphi}$, $x > 0$, and $\varphi \in (0, 1)$ is the damping constant used.

5. Simulation study

In this section we check what is the improvement of a forecasting accuracy depending on the intensity of time-close observations presence in data. This *intensity* means (1) the *frequency* of time-close observations in data and (2) the *closeness* of these time-close observations. In addition we try three levels of *smoothness* of the series. Totally we use 21 different simulated time series.

Firstly we generated a regular time series by a certain ARIMA(0, 2, 2) model (which is a reasonable theoretical model for time series treated by Holt method). Then we made a sampling from this time series by taking individual time steps of newly created irregular time series as realizations of a certain integer-valued random variable. Finally we normalized the time axis of the resulting irregular time series so that its average time spacing q is equal to 1.

We will use the more understandable parametrization of Holt method to describe the concrete settings of ARIMA(0, 2, 2) generating models. We always started with $S_0 = T_0 = 0$ and the one-step ahead forecasting errors were independent identically distributed with $N(0, 1)$. All generated irregular time series have 2000 observations.

The parameters of ARIMA(0, 2, 2) model, in Holt method parametrization, were taken as $\alpha^* = 1 - (1 - \alpha)^{1/Q}$ and $\gamma^* = 1 - (1 - \gamma)^{1/Q}$ where α and γ are 0.2 and 0.1 (low smoothness), 0.4 and 0.25 (medium smoothness) and 0.6 and 0.4 (high smoothness) respectively and Q is the expected value from Table 1, depending on the selected time step distribution.

Table 1. Used time step distributions. U is a discrete uniform distribution and δ_x is the Dirac measure at point x .

closeness	freq.	time step distribution	Q
-	-	$U\{1, 2, 3, 4\}$	2.5
low	low	$0.04 \delta_1 + 0.32 \delta_5 + 0.32 \delta_{10} + 0.32 \delta_{15}$	10
medium	low	$0.04 \delta_1 + 0.32 \delta_{10} + 0.32 \delta_{20} + 0.32 \delta_{30}$	19.6
high	low	$0.04 \delta_1 + 0.32 \delta_{20} + 0.32 \delta_{40} + 0.32 \delta_{60}$	38.8
low	high	$0.1 \delta_1 + 0.3 \delta_5 + 0.3 \delta_{10} + 0.3 \delta_{15}$	9.1
medium	high	$0.1 \delta_1 + 0.3 \delta_{10} + 0.3 \delta_{20} + 0.3 \delta_{30}$	18.1
high	high	$0.1 \delta_1 + 0.3 \delta_{20} + 0.3 \delta_{40} + 0.3 \delta_{60}$	36.1

As far as the implementation of both Wright’s original method and the modified one is concerned, the initial values of S and T have been constructed using the estimate of a linear regression through the first 10 observations of the series (discounted least squares with weights decreasing towards future with discount factor $1 - \sqrt{\alpha\gamma}$ have been used) and the minimal MSE values of smoothing constants α and γ have been used (do not mix with the generating parameters!). These optimal smoothing constants and the achieved minimal $RMSE = \sqrt{MSE}$ are presented for each of 21 time series and both the original and modified method, see Table 2.

From Table 2 we can see that the modified method has achieved a lower RMSE than the original one in all 21 cases. Also in all 21 cases the original method used higher α value and dramatically lower γ value than the modified one. This means that the original method tried to prevent a negative impact of time-close observations by choosing very low γ value and compensated this by choosing a higher α value. This difference between compared methods is more significant in cases of higher closeness and frequency of time-close observations. Generally, in contrast to the original method, the results from the modified one are not much dependent on which of 21 time series we take.

If we used $\gamma = 0.05$ or $\gamma = 0.1$ in cases where the original method has much lower values of γ as its optimum, we would usually obtain seriously wrong results characterized by high RMSE, high positive autocorrelation of forecasting errors (0.2–0.8) and high kurtosis of these errors (they have a heavy tailed sample distribution). Visual inspection would usually discover quite crazy patterns in forecasts: at a time of time-close observations the slope estimate is changed rapidly so the following forecasts are totally out of the mainstream of observations. Then it takes a certain time to join it again. Sometimes, since the forecasting errors during this returning period are huge, another time-close observations effect occurs which shoots the forecasts rapidly to the other direction. This can repeat several times and create a spurious oscillation in forecasts.

See Figure 1 for illustration of this phenomenon. There is a detail of one of 21 used simulated time series (that one with high frequency, high closeness and medium smoothness; it is the second one from bottom in Table 2). Forecasts obtained by both the methods are plotted, both using fixed $\alpha = 0.3$ and $\gamma = 0.1$ which is a usual expert guess combination of smoothing constant values. While the forecasts from the modified method are without problems, those from the original one oscillates crazily.

Figure 1. Forecasts obtained by the methods, both with fixed $\alpha = 0.3$ and $\gamma = 0.1$.

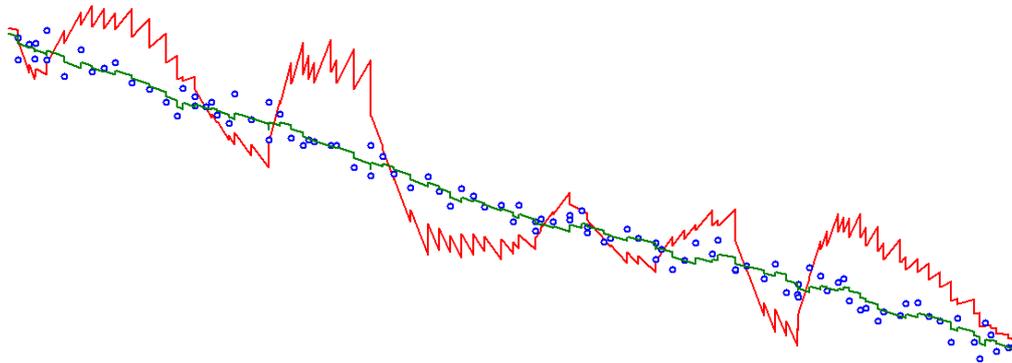


Table 2. Optimal α and γ values and the achieved RMSE for all 21 simulated time series, for both the original and the modified method.

freq.	closeness	smooth.	Original method			Modified method		
			α	γ	RMSE	α	γ	RMSE
-	-	low	0.1587	0.0494	1.0525	0.1559	0.0743	1.0503
-	-	medium	0.3791	0.1342	1.1129	0.3488	0.2006	1.0991
-	-	high	0.5326	0.2784	1.2202	0.5041	0.3589	1.1994
low	low	low	0.1304	0.0269	1.0654	0.1054	0.0599	1.0511
low	low	medium	0.3129	0.0684	1.1385	0.2412	0.1683	1.1035
low	low	high	0.4478	0.1473	1.2853	0.3781	0.2866	1.2256
low	medium	low	0.1461	0.0104	1.0798	0.1229	0.0500	1.0657
low	medium	medium	0.3295	0.0223	1.1690	0.2245	0.1277	1.1004
low	medium	high	0.4661	0.0585	1.3312	0.3467	0.2339	1.2187
low	high	low	0.1527	0.0059	1.0686	0.1092	0.0657	1.0411
low	high	medium	0.3268	0.0120	1.2137	0.2085	0.1089	1.0967
low	high	high	0.5319	0.0318	1.4497	0.3233	0.2145	1.2130
high	low	low	0.1376	0.0242	1.0242	0.1162	0.0738	1.0188
high	low	medium	0.3464	0.0536	1.1732	0.2647	0.1657	1.1311
high	low	high	0.5250	0.1043	1.3414	0.4482	0.2551	1.2495
high	medium	low	0.1662	0.0057	1.0859	0.1147	0.0591	1.0568
high	medium	medium	0.3229	0.0321	1.2119	0.2355	0.1441	1.1169
high	medium	high	0.5174	0.0456	1.4134	0.3576	0.2456	1.2243
high	high	low	0.1251	0.0044	1.0894	0.0865	0.0425	1.0498
high	high	medium	0.2666	0.0122	1.2148	0.2015	0.1223	1.1157
high	high	high	0.4989	0.0161	1.4289	0.3299	0.1900	1.2136

6. Conclusion

Modification of Wright’s version of Holt method for irregular time series suggested in this paper has proved to be a reasonable way to eliminate the impact of time-close observations. Only one formula needs to be modified by adding one term to it. Modified method has a better forecasting accuracy when compared with the original one by Wright. This improvement is just slight when the problem of time-close observations is not present and becomes substantial when time-close observations are present with higher intensity. Any special or extended version of Holt method (e.g. Holt method with damped trend, Holt-Winters method etc.) in its form for irregular time series can be improved just in the same way. There are no arguments against the usage of the suggested modification.

Acknowledgments. The paper is a part of research project MSM0021620839

References

- Cipra, T., Trujillo, J., Rubio, A., Holt-Winters method with missing observations, *Management Science*, 41, 174–178, 1995.
- Cipra, T., Exponential smoothing for irregular data, *Applications of Mathematics*, 51, 597–604, 2006.
- Cipra, T., Hanzák, T., Exponential smoothing for irregular time series, *Kybernetika*, 44, 385–399, 2008.
- Gardner, E. S., McKenzie, E., Forecasting trends in time series, *Management Science*, 31, 1237–1246, 1985.
- Wright, D. J., Forecasting data published at irregular time intervals using extension of Holt’s method, *Management Science*, 32, 499–510, 1986.
- Winters, P. R., Forecasting sales by exponentially weighted moving averages, *Management Science*, 6, 324–342, 1960.