KYBERNETIKA — VOLUME *48* (2012), NUMBER 1, PAGES 1–15

HOLT–WINTERS METHOD WITH GENERAL SEASONALITY

Tomáš Hanzák

The paper suggests a generalization of widely used Holt–Winters smoothing and forecasting method for seasonal time series. The general concept of seasonality modeling is introduced both for the additive and multiplicative case. Several special cases are discussed, including a linear interpolation of seasonal indices and a usage of trigonometric functions. Both methods are fully applicable for time series with irregularly observed data (just the special case of missing observations was covered up to now). Moreover, they sometimes outperform the classical Holt–Winters method even for regular time series. A simulation study and real data examples compare the suggested methods with the classical one.

Keywords: exponential smoothing, Holt–Winters method, irregular time series, seasonal indices, trigonometric functions

Classification: 62M10, 62M20, 60G35, 65D10

1. INTRODUCTION

Holt–Winters method employs p seasonal indices (additive or multiplicative) to model the seasonal pattern of length p, see e. g. [8] or [14]. Many modifications appeared later, introducing different trend types than the basic (locally) linear: exponential, damped linear or damped exponential, see [6] for a nice overview.

However, smaller attention was paid to the seasonality modeling, even though the usage of seasonal indices brings couple of limitations, see Section 2. Holt–Winters method with general seasonality modeling is therefore suggested in this paper. The goal is to offer a broader spectrum of possibilities for seasonality treatment while staying in the widely known and understood framework of the Holt–Winters method. The suggested methods are applicable also for irregular time series and both additive and multiplicative seasonality is offered.

Model based approach to exponential smoothing (various ARIMA, SARIMA and state space models) was often applied, see e. g. [1, 10, 11] or [3]. In contrast to that, the method suggested in this paper can be viewed as ad-hoc, following the tradition of exponential weighting idea from [4, 8, 14, 15] or [7]. This hopefully supports the understandability of the method while it does not harm its smoothing and forecasting performance; [1] and [3] showed that the performance of ad-hoc methods is fairly comparable with that of the optimal model based ones.

In Section 2 the classical Holt–Winters method is reminded. In Section 3 Holt–Winters method with a general seasonality modeling (in its additive and multiplicative variants) is presented. The properties of the method are discussed, its theoretical justification based on *Discounted Least Squares* (DLS) estimation is given and the implementation details are outlined here. In Section 4 we discuss particular methods useful in practice, including linearly interpolated seasonal indices and trigonometric functions. Sections 5 and 6 compare the suggested methods numerically with the classical one on simulated and real data, respectively. Section 7 brings the summary of the paper.

2. CLASSICAL HOLT–WINTERS METHOD

To remind it and to unify its notation, we release the basic formulas of the classical Holt–Winters method here. Let $\{y_t, t \in \mathbb{Z}\}$ be a *regular* time series with *locally linear trend* and *additive seasonality* of period $p \geq 2$. We consider its *level* L_t, *slope* T_t and *seasonal index* S_t at time t. S's are supposed to approximately sum up to 0 and repeat after period p.

The forecast $\hat{y}_{t+\tau}(t)$ of the future unknown observation $y_{t+\tau}$, $\tau > 0$, constructed at time t, is

$$\hat{y}_{t+\tau}(t) = L_t + \tau \cdot T_t + S_{t\oplus\tau}\,, \tag{1}$$

where $t\oplus\tau = t+1-p+[(\tau-1) \mod p]$. After a new observation y_{t+1} becomes available, the level, slope and seasonal index are updated using the recursive formulas

$$\begin{aligned} L_{t+1} &= (1-\alpha)\cdot(L_t + T_t) + \alpha\cdot(y_{t+1} - S_{t+1-p})\,, && (2)\\ T_{t+1} &= (1-\gamma)\cdot T_t + \gamma\cdot(L_{t+1} - L_t)\,, && (3)\\ S_{t+1} &= (1-\delta)\cdot S_{t+1-p} + \delta\cdot(y_{t+1} - L_{t+1})\,, && (4) \end{aligned}$$

where $\alpha, \gamma, \delta \in (0, 1]$ are *smoothing constants* (for level, slope and seasonal indices). Equations (2) – (4) are often rewritten to their equivalent *error-correction* form (see e. g. [5] or [6]):

$$\begin{aligned} L_{t+1} &= L_t + T_t + \alpha\cdot e_{t+1}\,, && (5)\\ T_{t+1} &= T_t + \alpha\cdot\gamma\cdot e_{t+1}\,, && (6)\\ S_{t+1} &= S_{t+1-p} + (1-\alpha)\cdot\delta\cdot e_{t+1}\,, && (7) \end{aligned}$$

where $e_{t+1} = y_{t+1} - \hat{y}_{t+1}(t)$ is the one-step-ahead forecasting error at time $t+1$.

To use the seasonal indices, we must be able to assign each observation to exactly one of p calendar units forming the complete period (e. g. January, February etc. for monthly observations with annual seasonality, $p = 12$). This is still possible in a time series with missing observations, see [4] for such an extension of Holt–Winters method. However, the calendar assignment is not possible in a general irregular time series and so there was no Holt–Winters method available for this case up to now.

Time series with $p \gg 0$, i. e. with many observations per one period, are also not favorable for the classical Holt–Winter method since we need to carry out enormous number of seasonal indices to form the seasonal pattern. This is unpleasant especially when the seasonal pattern is relatively smooth.

The above mentioned issues can be overcome by using a different or extended seasonality modeling while not leaving the widely understandable framework of Holt–Winters method at the same time. The general approach and its special cases are presented in Sections 3 and 4.

3. GENERAL SEASONALITY MODELING IN HOLT–WINTERS METHOD

Seasonality can be generally modeled using $K \geq 1$ different real-valued functions $f_1, f_2, \ldots, f_K$, all defined on $\mathbb{R}$. Each f_k is supposed to be periodic with a specific period $p_k \in (0, +\infty)$. The seasonal pattern S is formed as a linear combination of f_k as in a linear regression:

$$S(t) = \sum\nolimits_{k=1}^{K} A^k \cdot f_k(t)\,, \tag{8}$$

where $t \in \mathbb{R}$ is time and $A^k \in \mathbb{R}$ are appropriate *amplitudes*.

In the case of an *additive* seasonality, this $S(t)$ is then added to the time series level L_t to create the smoothed value:

$$\hat{y}_t = L_t + S(t) \tag{9}$$

while to get a *multiplicative* seasonality, L_t is multiplied by the exponential of $S(t)$:

$$\hat{y}_t = L_t \cdot \exp[S(t)]\,. \tag{10}$$

We suppose f_k just to be bounded. One can take functions f_k centered to 0 in a certain sense. It is also reasonable (but not necessary) for f_k to be linearly independent, see (8).

3.1. Method formulation

Now we will incorporate the above described general seasonality modeling concept into Holt–Winters method. Let $\{y_{t_n}, n \in \mathbb{Z}\}$, $t_{n+1} > t_n$, be an *irregular* seasonal time series with locally linear trend (the other trend types can be used as well) and additive seasonality (multiplicative case will be described later in Section 3.3). We consider its *level* L_{t_n}, *slope* T_{t_n} and *seasonal component*

$$S_{t_n}(t) = \sum\nolimits_{k=1}^{K} A^k_{t_n} \cdot f_k(t) \tag{11}$$

at time t_n. Here $A^k_{t_n}$ are adaptive amplitudes valid at time t_n. We must correctly distinguish between the two different times t and t_n here.

The forecast $\hat{y}_{t_n+\tau}(t_n)$ and the smoothed value $\hat{y}_{t_n}$ are analogous to (1):

$$\hat{y}_{t_n+\tau}(t_n) \;=\; L_{t_n} + \tau \cdot T_{t_n} + S_{t_n}(t_n + \tau)\,, \tag{12}$$

$$\hat{y}_{t_n} \;=\; L_{t_n} + S_{t_n}(t_n)\,. \tag{13}$$

After a new observation $y_{t_{n+1}}$ becomes available, the level L, slope T and the K seasonal amplitudes A^k, $k = 1, \ldots, K$, are updated using error-correction formulas analogous to (5) – (7):

$$L_{t_{n+1}} \;=\; L_{t_n} + (t_{n+1} - t_n)\, T_{t_n} + \alpha_{t_{n+1}} e_{t_{n+1}}\,, \tag{14}$$

$$T_{t_{n+1}} \;=\; T_{t_n} + \alpha_{t_{n+1}} \gamma_{t_{n+1}} e_{t_{n+1}} / (t_{n+1} - t_n)\,, \tag{15}$$

$$A^k_{t_{n+1}} \;=\; A^k_{t_n} + (1 - \alpha_{t_{n+1}})\, \delta^k_{t_{n+1}} e_{t_{n+1}} / f_k(t_{n+1})\,, \tag{16}$$

where $e_{t_{n+1}} = y_{t_{n+1}} - \hat{y}_{t_{n+1}}(t_n)$ and we take $0/0 = 0$ by definition in (16). Formulas (14) and (15) are equivalent to those in [15] and [4]. The factor $(1 - \alpha_{t_{n+1}})\delta^k_{t_{n+1}}$ in (16) expresses the portion of $e_{t_{n+1}}$ which is absorbed to the kth seasonal component $A^k_{t_{n+1}} f_k(t_{n+1})$. The division by $f_k(t_{n+1})$ in (16) is due to (11) and it is not in conflict with the *additive* seasonality used.

Smoothing coefficient $\alpha_{t_n} \in (0, 1]$ for level in (14) is updated in a recursive way, following the basic idea of exponential weighting, exactly as in [15] and [4]:

$$\alpha_{t_{n+1}} = \frac{\alpha_{t_n}}{\alpha_{t_n} + (1-\alpha)^{t_{n+1}-t_n}}\,, \tag{17}$$

where $\alpha \in (0, 1]$ is a smoothing constant for level.

For the smoothing coefficient $\gamma_{t_n} \in (0, 1]$ for slope in (15), we will rather use a modified updating formula from [7]:

$$\gamma_{t_{n+1}} = \frac{\gamma_{t_n}}{\gamma_{t_n} + \frac{t_n - t_{n-1}}{t_{n+1}-t_n}(1-\gamma)^{t_{n+1}-t_n}}\,, \tag{18}$$

where $\gamma \in (0, 1]$ is a smoothing constant for slope. For irregular time series, this differs from that one used in [15] or [4]:

$$\gamma^*_{t_{n+1}} = \frac{\gamma^*_{t_n}}{\gamma^*_{t_n} + (1-\gamma)^{t_{n+1}-t_n}}\,. \tag{19}$$

The modified coefficient $\gamma_{t_{n+1}}$ defined by (18) makes the slope estimate $T_{t_{n+1}}$ in (15) safe from a negative impact of the time distance $t_{n+1} - t_n$ being close to zero. See [7] for further details and evaluation of this modification on the non-seasonal Holt method.

Smoothing coefficients $\delta^k_{t_n}$, $k = 1, \ldots, K$, for the seasonal amplitudes in (16) are also updated in a recursive way. We consider K generally different smoothing constants $\delta^k \in (0, 1]$ belonging to each of the functions f_k (but we can take $\delta_k \equiv \delta$ as a special case). For $k = 1, \ldots, K$ let us denote

$$W^k_{t_n} \equiv \sum\nolimits_{j=0}^{+\infty} \left(1 - \delta^k\right)^{t_n - t_{n-j}} f_k^2(t_{n-j})\,. \tag{20}$$

Obviously W^k can be easily updated recursively over time:

$$W^k_{t_{n+1}} = \left(1 - \delta^k\right)^{t_{n+1}-t_n} \cdot W^k_{t_n} + f_k^2(t_{n+1})\,. \tag{21}$$

For $k = 1, \ldots, K$, let us further denote the dimensionless quantities

$$\Delta^k_{t_{n+1}} \equiv f_k^2(t_{n+1}) \,/\, W^k_{t_{n+1}} \tag{22}$$

(we take again $0/0 = 0$). Since according to (20) it is $0 \le f_k^2(t_{n+1}) \le W^k_{t_{n+1}}$, we have $\Delta^k_{t_{n+1}} \in [0, 1]$. This is declared to be the ideal value for $\delta^k_{t_{n+1}}$ in the case that $K = 1$, i. e. if there was no competition between individual f_k's.

Formula (22) is consistent with the fundamental idea of exponential weighting, see [15] for simple exponential smoothing. In (20) together with (22), besides the observation

time t_{n-j}, we measure the relevance of a particular observation $y_{t_{n-j}}$ with respect to A^k also by the magnitude $f_k^2(t_{n-j})$. This expresses the fact that if $f_k(t_{n-j}) \approx 0$ then the observation at time t_{n-j} contains very little information about the value of A^k. In Section 3.2 we give additional justification for the choice in (20) and (22).

However, if $K > 1$ (which is typically the case), it can happen that $\sum_{k=1}^{K} \Delta_{t_{n+1}}^k > 1$ which implies that the total portion of the error absorbed would exceed 100 % if one used $\delta_{t_{n+1}}^k = \Delta_{t_{n+1}}^k$. So it is necessary to normalize $\Delta_{t_{n+1}}^k$ in a suitable way to get the final coefficients $\delta_{t_{n+1}}^k$. We let

$$\Delta_{t_{n+1}} \equiv 1 - \prod_{k=1}^{K} \left(1 - \Delta_{t_{n+1}}^k\right) \in [0,1] \tag{23}$$

to be the total portion of the error absorbed instead of

$$D_{t_{n+1}} \equiv \sum_{k=1}^{K} \Delta_{t_{n+1}}^k \geq 0\,. \tag{24}$$

To achieve this, let us take the final smoothing coefficients $\delta_{t_{n+1}}^k$ as

$$\delta_{t_{n+1}}^k \equiv \frac{\Delta_{t_{n+1}}}{D_{t_{n+1}}} \cdot \Delta_{t_{n+1}}^k \in [0,1]\,,\ \ k = 1, \ldots, K \tag{25}$$

(again take $0/0 = 0$). The motivating interpretation of (23) vs. (24) is that we rather imagine independence than disjointness of the k events of absorbtion with probabilities $\Delta_{t_{n+1}}^k$.

Let us summarize that the suggested Holt–Winters method with general seasonality consists of formulas (14) – (18) and (21) – (25). One needs to keep totally $4 + 2K$ numerical variables in memory which are updated through the time by the above listed recursive formulas. The computational complexity of the method is comparable with that from [4] and is reduced with lower number K of seasonal functions f_k or when some of them are repeatedly equal to 0 (see Section 4 for concrete examples).

3.2. Properties of the method and its theoretical justification

The smoothing coefficients $\delta_{t_{n+1}}^k$ as defined in (20) – (25) have reasonable properties:

- By $A_{t_{n+1}}^k$ update we move from $\hat{y}_{t_{n+1}}(t_n)$ closer to $y_{t_{n+1}}$. Summing the k movements, we come to $S_{t_{n+1}}(t_{n+1}) = S_{t_n}(t_{n+1}) + (1 - \alpha_{t_{n+1}})\Delta_{t_{n+1}} e_{t_{n+1}}$, see (11), (16), (24), (25). So the total portion of $e_{t_{n+1}}$ absorbed by seasonals is $(1-\alpha_{t_{n+1}})\Delta_{t_{n+1}} \in [0,1]$. Compare this with (7).

- The error $e_{t_{n+1}}$ is absorbed more to f_k with higher δ^k (i. e. it has really the meaning of a smoothing constant) and with $f_k^2(t_{n+1})$ larger compared to its recent values, see (20) and (22).

- If $f_k(t_{n+1}) \to 0$ then (ceteris paribus) $\delta_{t_{n+1}}^k / f_k(t_{n+1}) \to 0$. This means that we do not need to worry about values of f_k near to 0, see the division in (16).

To justify the concrete choice in (20) and (22) for $\delta^k \equiv \delta$, let us consider a *Discounted Least Squares* (DLS) estimation of K parameters A^k in the linear regression model

$$y_t \approx \sum\nolimits_{k=1}^{K} A^k \, f_k(t) \tag{26}$$

with discount factor $1-\delta \in (0,1)$. The minimized criterion based on the data up to time t_n is

$$\Sigma_n(\boldsymbol{A}) \equiv \sum_{j=0}^{\infty} \left[y_{t_{n-j}} - \sum\nolimits_{k=1}^{K} A^k \, f_k(t_{n-j}) \right]^2 (1-\delta)^{t_n - t_{n-j}} \,, \tag{27}$$

where we denoted $\boldsymbol{A} = (A^1, \ldots, A^K)'$. We purposely do not consider the level-trend component $L + t \cdot T$ in (26) since we focus on the seasonal smoothing coefficients δ_{t_n} here (we can think of y here as being after a trend elimination).

Denote by $\boldsymbol{A}_{t_n}$ the argument of minima of $\Sigma_n(\boldsymbol{A})$. It is

$$\boldsymbol{A}_{t_n} = (\boldsymbol{F}'_n \boldsymbol{D}_n \boldsymbol{F}_n)^{-1} \boldsymbol{F}'_n \boldsymbol{D}_n \boldsymbol{Y}_n \,, \tag{28}$$

where $\boldsymbol{F}_n = \{f_k(t_{n-j})\}_{j=0,1,2,\ldots}^{k=1,\ldots,K}$ is the regression design matrix, $\boldsymbol{D}_n = \mathrm{Diag}\{1, (1-\delta)^{t_n - t_{n-1}}, (1-\delta)^{t_n - t_{n-2}}, \ldots\}$ is the diagonal discounting matrix and $\boldsymbol{Y}_n = (y_{t_n}, y_{t_{n-1}}, y_{t_{n-2}}, \ldots)'$[1]. Further denote

$$\hat{y}_{t_{n+1}}(t_n) = \sum\nolimits_{k=1}^{K} A^k_{t_n} \, f_k(t_{n+1}) \tag{29}$$

the regression prediction of $y_{t_{n+1}}$ using the estimate $\boldsymbol{A}_{t_n}$ and $e_{t_{n+1}} = y_{t_{n+1}} - \hat{y}_{t_{n+1}}(t_n)$ the corresponding prediction error. Since it is

$$\Sigma_{n+1}(\boldsymbol{A}) = (1-\delta)^{t_{n+1} - t_n} \cdot \Sigma_n(\boldsymbol{A}) + \left[y_{t_{n+1}} - \hat{y}_{t_{n+1}}(t_{n+1}) \right]^2 \,, \tag{30}$$

$y_{t_{n+1}} = \hat{y}_{t_{n+1}}(t_n)$ implies $\boldsymbol{A}_{t_{n+1}} = \boldsymbol{A}_{t_n}$[2]. This fact together with (28) gives us

$$\boldsymbol{A}_{t_{n+1}} = \boldsymbol{A}_{t_n} + (\boldsymbol{F}'_{n+1} \boldsymbol{D}_{n+1} \boldsymbol{F}_{n+1})^{-1} \{f_k(t_{n+1})\}_{k=1,\ldots,K} \cdot e_{t_{n+1}} \,, \tag{31}$$

where $\{f_k(t_{n+1})\}_{k=1,\ldots,K}$ is the first column ($j=0$) of matrix $\boldsymbol{F}'_{n+1}$.

Given that $K \times K$ matrix $\boldsymbol{F}'_{n+1} \boldsymbol{D}_{n+1} \boldsymbol{F}_{n+1}$ is diagonal (i. e. the regressors f_k are orthogonal in the sense that $f_k \cdot f_l \equiv \sum_{j=0}^{+\infty} (1-\delta)^{t_{n+1} - t_{n+1-j}} f_k(t_{n+1-j}) f_l(t_{n+1-j}) = 0$ for all $k \neq l$), we get

$$A^k_{t_{n+1}} = A^k_{t_n} + \frac{f_k(t_{n+1})}{W^k_{t_{n+1}}} e_{t_{n+1}} = A^k_{t_n} + \Delta^k_{t_{n+1}} e_{t_{n+1}} / f_k(t_{n+1}) \,, \tag{32}$$

where $W^k_{t_{n+1}}$ and $\Delta^k_{t_{n+1}}$ are defined exactly as before. This result support the definition of $\Delta^k_{t_{n+1}}$ in (22).

[1] The infinite dimension of the matrices $\boldsymbol{F}_n$, $\boldsymbol{D}_n$ and $\boldsymbol{Y}_n$ just turns the scalar products from finite sums to series sums convergent due to exponential decay of $(1-\delta)^{t_n - t_{n-j}}$.

[2] $\boldsymbol{A}_{t_n}$ makes $\Sigma_n(\boldsymbol{A})$ minimal and the second summand 0 due to $\hat{y}_{t_{n+1}}(t_{n+1}) = \hat{y}_{t_{n+1}}(t_n) = y_{t_{n+1}}$.

Ignoring the possible non-zero off-diagonal elements of matrix $\boldsymbol{F}'_{n+1}\boldsymbol{D}_{n+1}\boldsymbol{A}_{n+1}$ is the reason why we need to do the normalization in (25). If we solved correctly $K \times K$ matrix inversion in (31), we would receive directly reasonable values for $\delta^k_{t_{n+1}}$ with no additional normalization needed.

Since K is typically quite large (e. g. 12), we prefer the simplified approach of (20) – (25) based on the diagonality assumption. If the functions f_k are approximately orthogonal (i. e. their scalar products $f_k \cdot f_l$ for $k \neq l$ are *almost* zero when compared to $f_k \cdot f_k = W^k_{t_{n+1}}$) then this is an acceptable approximation.

Another possible approach, not using the ideas of Holt–Winters method at all, would be to regress y on the regressors $\{1, t, f_1(t), \ldots, f_K(t)\}$ using DLS estimation method with a certain discount factor. But besides facing the necessity of inverting $(K+2) \times (K+2)$ matrices, we lose the important flexibility of having three independent smoothing constants as in Holt–Winters method.

3.3. Multiplicative seasonality

Up to now we have considered only the case of an *additive* seasonality. To get a *multiplicative* seasonality, one has to replace the additive prediction and smoothing formulas (12) and (13) with

$$\hat{y}_{t_n+\tau}(t_n) = (L_{t_n} + \tau \cdot T_{t_n}) \cdot \exp\left[S_{t_n}(t_n + \tau)\right] , \tag{33}$$

$$\hat{y}_{t_n} = L_{t_n} \cdot \exp\left[S_{t_n}(t_n)\right] . \tag{34}$$

The recursive formula (16) for the amplitudes update is simply changed to

$$A^k_{t_{n+1}} = A^k_{t_n} + (1 - \alpha_{t_{n+1}})\, \delta^k_{t_{n+1}} \left[\ln y_{t_{n+1}} - \ln \hat{y}_{t_{n+1}}(t_n)\right] . \tag{35}$$

By taking the natural logarithm of the multiplicative forecasting error $y_{t_{n+1}}/\hat{y}_{t_{n+1}}(t_n)$ we simply convert it from the multiplicative world of y to the additive world of f_k and A^k. In (33) and (34) we do the reverse conversion from additive to multiplicative.

So just the exponential and logarithm transformations must be placed correctly into formulas (12), (13) and (16) of the additive method to switch completely to the multiplicative seasonality. This enables us to implement the both variants as a single piece of programm code.

3.4. Practical implementation

To apply successfully the above described smoothing and forecasting method, one must necessarily deal with the following tasks:

- To choose suitable seasonality modeling functions f_k, specially their number K, depending on the nature of the seasonal pattern. The standardized choices are suggested in Section 4. Generally with higher K we are able to model more precisely even complicated patterns but we must beware of over-fitting. See Sections 5 and 6 for practical experiences.

- To choose the values of $K + 2$ smoothing constants α, γ and δ^k, $k = 1, \ldots, K$. It seems reasonable to reduce the number of parameters by taking $\delta^k \equiv \delta$. The three constants α, γ and δ can be searched numerically over the unit cube $(0, 1]^3$.

- To set up the initial values L_0, T_0, α_0, δ_0, A_0^k and W_0^k before running the recursive computation. We recommend using the general approach of *backcasting* (*backward forecasting*, see [2] for a brief explanation). To initialize the backcasting itself we can put simply $A_0^k = 0$ and W_0^k based on a rough approximation of (20) (let $\overline{f_k^2}$ be the average squared value of f_k over the available observation times and q the average time spacing of the series):

$$W_0^k \approx \sum_{j=0}^{+\infty} (1-\delta^k)^{jq}\, \overline{f_k^2} = \frac{\overline{f_k^2}}{1-(1-\delta^k)^q}\,. \tag{36}$$

4. USEFUL SPECIAL CASES

4.1. Classical Holt–Winters method

To get the classical Holt–Winters method (for regular time series) with p seasonal indices and period $p \geq 2$, see (1) – (7), we simply take $K = p$ and

$$f_k(t) = \begin{cases} 1 & \text{if } (t \mod p) = k\,, \\ 0 & \text{otherwise}\,. \end{cases} \tag{37}$$

So f_k are the indicators of individual calendar units, it is $p_k \equiv p$ and f_k are perfectly orthogonal (it is even $f_k(t) f_l(t) = 0$ for all $k \neq l$ and $t \in \mathbb{R}$). Further take $\delta^k \equiv \delta$. The seasonal smoothing coefficients are of a trivial form:

$$\delta_t^k = \begin{cases} 1-(1-\delta)^p & \text{if } (t \mod p) = k\,, \\ 0 & \text{otherwise}\,. \end{cases} \tag{38}$$

So only one amplitude A^k (belonging to the actual calendar unit of t) is updated in one time step, the remaining ones stay unchanged. It is $\overline{f_k^2} = 1/p$. Notice that $\delta_t^k = 1-(1-\delta)^p \neq \delta$ due to the p time steps between the two consecutive observations from the same calendar unit. But this is just a different parametrization of the method.

However, for multiplicative seasonality, we get a slightly different smoothing formulas for the seasonal indices. The classical method *additively* averages the old and the new values of the seasonal index while our method does this mixing in terms of a weighted *geometric* mean. This multiplicative treating of multiplicative seasonal indices seems to be more reasonable and consistent.

4.2. Normalized seasonal indices

In [2] the possibility to normalize the seasonal indices in Holt–Winters method to ensure that they always sum up to 0 is mentioned. This is a reasonable normalizing condition which helps us to strictly separate the level and the seasonal component. We can employ such a normalizing in our general seasonality concept. Just replace (37) with

$$f_k(t) = \begin{cases} 1 & \text{if } (t \mod p) = k\,, \\ -1/(p-1) & \text{otherwise}\,. \end{cases} \tag{39}$$

Functions $f_k(t)$ are now centered to 0 and so the whole seasonal component $S(t)$ defined in (11) is as well. Functions $f_k(t)$ are still linearly independent and approximately

orthogonal for $p \gg 0$. Now always all the amplitudes A^k are updated in a single time step.

4.3. Missing observations

By taking K, f_k and δ^k the same as in Section 4.1 and just allowing the analyzed time series y to have missing observations (so the calendar assignment is still possible), we come to the method from [4]. Again only the one amplitude A^k belonging to the actual calendar unit of t is updated in a single time step. But now the non-zero smoothing coefficient δ_t^k varies step by step, depending on the value of W^k which contains the information about the time structure of the series when the current calendar unit is concerned.

4.4. Interpolated seasonal indices

To cover the inter-calendar observations or to reduce the number of seasonal indices used, it is possible to interpolate linearly the neighboring indices. We will describe this directly for the number $K \geq 2$ of the seasonal indices used independent of the period length $p \in (0, +\infty)$ and with the general time axis origin $o \in \mathbb{R}$. Let us define

$$f_k(t) = \left\{ 1 - \min_{j \in \mathbb{Z}} \left| \frac{K \cdot (t-o)}{p} - (j \cdot K + k) \right| \right\}^+ . \tag{40}$$

Each of f_k has the form of p-periodic sequence of identical isosceles triangles with the basis length of $2p/K$ and the height of 1. The neighboring f_k's are shifted by p/K to each other and it is $f_K(o) = 1$. By setting the amplitudes A^k we can form any p-periodic $K-$piecewise linear seasonal pattern. The amplitudes A^k will appear as the pattern values at the K equidistant break points.

Functions $f_k(t)$ are still linearly independent. But they are not perfectly orthogonal since the neighboring triangles always overlap by one half of their bases. We can take routinely $\overline{f_k^2} = 1/K$ or $\overline{f_k^2} = 2/3 \cdot 1/K$ depending on the layout of the observation times. In a single time step, only one (if $f_k(t) = 1$ for some k) or two (otherwise) amplitudes are updated. In the later case, the two updated amplitudes belong to the two indices surrounding the observation time t.

The practical choice of K and o should reflect the smoothness of the seasonal pattern and the layout of observation times. For example, the following three settings can be tested in practice for regular time series with period $p \in \mathbb{N}$ (see Sections 5 and 6):

- **Classical method:** $K = p$ and $o = 0$. It is an extension of the method for missing observations from [4].
- **Shifted seasonal indices:** $K = p$ and $o = 0.5$. All the observations are shifted by 0.5 so they are all treated as inter-calendar. Always the two surrounding indices are composed (by their simple arithmetic average) to form the corresponding seasonal component.
- **Sparse seasonal indices:** $K = p/2$ together with $o = 0$ or $o = 1$. This is suitable for large p and relatively smooth seasonal pattern.

Of course we must beware of interpolating observation at peak or low point of the seasonal pattern.

4.5. Trigonometric functions

As an alternative to the seasonal indices, we can use trigonometric functions of time to model the seasonality. The seasonal pattern will be composed from several harmonic curves of different periods. Since usually both the amplitude and phase of the harmonic curve are unknown (and/or variable during time) we will always involve sine and cosine functions of the same period. The individual periods p_k will be taken as $p, p/2, p/3, \ldots$ where p is the period length of the series. For example, when $K = 4$ (only even values of K are used), we define

$$f_1(t) = \sin\frac{2\pi t}{p}\,,\quad f_2(t) = \cos\frac{2\pi t}{p}\,,\quad f_3(t) = \sin\frac{4\pi t}{p}\,,\quad f_4(t) = \cos\frac{4\pi t}{p}\,. \tag{41}$$

The user just has to specify the value of $h = K/2$, i. e. the number of full harmonics to be included. Sometimes even $h = 1$ can give good results, the values $h = 2, 3$ or 4 are applicable in most cases. It should always be $2h \le p/q$ to prevent from over-fitting.

Let us notice that the trigonometric functions f_k are centered to 0, linearly independent and approximately orthogonal (exact orthogonality holds for $\delta = 0$ and defining the scalar product in a continuous way as $f_k \cdot f_l = \int_0^p f_k(t) f_l(t) \mathrm{d}t$). Since the sine and cosine functions are equal to 0 only at isolated time points, usually all the seasonal amplitudes are updated in each time step. Since $\sin^2 t + \cos^2 t \equiv 1$, we can take routinely $\overline{f_k^2} \equiv 1/2$.

4.6. Multiple seasonality

Half hourly electricity demand time series contains two different seasonalities: daily (period 48) and weekly (period $7 \cdot 48 = 336$). To make forecasts, [12] used *double seasonal* Holt–Winters method with two sets of seasonal indices (48 and 336 indices for the daily and weekly seasonality, respectively). Another application of such methods can found e. g. in [13].

Such a multiple seasonality can be obtained as a special case of our general concept. We simply take two sets of indicator functions f_k as in (37), with $p_k = 48$ for the daily set and $p_k = 336$ for the weekly one.

5. SIMULATION STUDY

In this section we will test the classical Holt–Winters method (Section 4.1), the method with shifted seasonal indices (Section 4.4) and the method with trigonometric functions (Section 4.5) on the simulated regular time series with locally constant trend and additive seasonality with period length $p = 7, 12$ and 24. The generating model used is

$$y_t = L_t + S_t + \varepsilon_t\,,\quad \varepsilon_t \sim iid\ N(0,1)\,, \tag{42}$$

$$L_t = L_{t-1} + \mu_t\,,\quad \mu_t \sim iid\ N(0, 0.1^2) \tag{43}$$

with ε_t and μ_t mutually independent. The seasonal component S_t used in (42) follows

$$S_t = (1-\nu)\cdot(S_{t-1} + S_{t-p} - S_{t-p-1}) - \nu \cdot \sum\nolimits_{j=t-p+1}^{t-1} S_j + \pi_t\,,\quad \pi_t \sim iid\ N(0,1)\,, \tag{44}$$

i. e. a special AR($p+1$) process. Seasonal innovations $\{\pi_t\}$ are independent of $\{\mu_t\}$ and $\{\varepsilon_t\}$. The parameter $\nu \in [0,1]$ rules the normalization of S to sum up to 0 and the smoothness of the seasonal pattern (lower ν creates a smoother pattern). We initialize (42) – (44) by $L_0 = 0$ and $S_j = 0$ for $j = -p, \ldots, 0$.

In SARIMA or state space models (see e. g. [10, 11]) the seasonal component for each calendar unit usually follows a random walk (i. e. the whole $\{S_t\}$ follows AR(p) process $S_t = S_{t-p} + \pi_t$). This means that the seasonal indices for different calendar units are independent and the formed seasonal pattern is not autocorrelated or smooth at all. But such a situation is rather rare in reality and thus the model (44) is more realistic in our opinion.

For a given p, we simulate time series of length $21p$, i. e. 21 complete periods. The first 10 periods are thrown away to eliminate the impact of initialization by $S \equiv 0$ in (44). The next 10 periods are used to initialize the methods and to optimize the smoothing constants α and δ in order to minimize RMSE (*Root Mean Square Error*) of one-step-ahead forecasting errors (we use $\gamma = 0.05$ fixed). The number h of full harmonics is also optimized when needed. We try $h = 2, 3$ for $p = 7$, $h = 2, 3, 4, 5$ for $p = 12$ and $h = 4, 5, 6, 7, 8$ for $p = 24$.

The last period is used to evaluate the out-of-sample forecasting accuracy. We calculate RMSE from all the possible combinations of forecasting times from $20p$ to $21p - 1$ and forecasting horizons from 1 to p, i. e. from $p(p+1)/2$ forecasting errors totally.

For each p we use $\nu = 0.05, 0.1$ and 0.2 in (44) and for each combination of p and ν we simulate 100 time series. This means that totally 5100-times the constants α and δ are optimized. We use a locally constant trend (instead of locally linear), see (42) and (43), and a low fixed value of $\gamma = 0.05$ purposely to prevent from three-dimensional smoothing constants optimization. All the computations were implemented in a specialized application developed by the author (the same holds for Section 6).

		Classical H-W		Shifted indices		Trigonometric	
p	ν	RMSE	ranking	RMSE	ranking	RMSE	ranking
7	0.05	2.910	1.88	2.786	1.94	2.932	2.17
7	0.1	2.761	1.91	2.624	1.85	2.785	2.24
7	0.2	2.195	2.01	2.185	2.17	2.119	1.82
12	0.05	3.626	2.11	3.347	1.77	3.579	2.12
12	0.1	3.551	2.04	3.180	1.95	3.120	2.01
12	0.2	2.291	1.86	2.243	1.80	2.361	2.34
24	0.05	7.294	2.21	3.246	1.45	4.402	2.34
24	0.1	4.023	2.12	2.870	1.46	3.788	2.42
24	0.2	2.189	1.48	2.222	1.76	2.635	2.76

Tab. 1. Average out-of-sample RMSE and average ranking of the three methods tested.

The average out-of-sample RMSE and the average ranking of the methods (1 = best, 3 = worst) are presented in Table 1. All the three methods seem to be relevant competitors and can be recommended for testing in practice. The "Shifted indices" method is

the best one in our simulation in most cases. However, also the classical Holt–Winters method and the method with trigonometric functions generally work reasonably. The results surprisingly do not depend much on the parameter ν (except the case of $p = 24$).

One must beware that the results of the simulation study are probably far determined by the particular generating model for the seasonal component, see (44). It is easy to generate time series for which the particular method is optimal and to illustrate the lack of performance of the remaining ones. But it is non-trivial to set up a neutral generating model useful for the comparison of the methods.

6. REAL DATA EXAMPLES

Now we will illustrate the methods on real time series data. For this purpose, we have downloaded five regular monthly time series (i. e. containing annual seasonality, $p = 12$) from [9]:

1. **AIR** – Int. airline passengers, monthly totals in thousands, 1949-1960 (144 observations);
2. **TEMP** – New York City monthly average temperatures, 1946-1959 (168 observations);
3. **GAS** – Monthly residential gas usage in Iowa, 1971-1979 (106 observations);
4. **LEVEL** – Lake Erie, monthly levels, 1921-1970 (600 observations);
5. **FLOW** – Tree River, mean monthly flows, 1969-1976 (96 observations).

In addition to Section 5, we will test also the "Sparse indices" method from Section 4.4. For **AIR**, **GAS** and **FLOW** we use a multiplicative seasonality, for **TEMP** and **LEVEL** the additive seasonality is used in all four methods.

The smoothing constants α, γ and δ and (if needed) the number h of full harmonics are optimized with respect to RMSE based on one-step-ahead forecasting errors through the whole series. The same in-sample RMSE values are reported in Table 2, together with the sample first order autocorrelation coefficients ρ_e of the forecasting errors. The table also contains the optimal value of h for each series.

	Classical H-W		Shifted indices		Sparse indices		Trigonometric		
Series	RMSE	ρ_e	RMSE	ρ_e	RMSE	ρ_e	h	RMSE	ρ_e
AIR	10.69	.237	10.25	-.124	16.44	-.126	5	10.41	.203
TEMP	0.740	.180	0.693	.114	0.799	-.181	1	0.713	-.121
GAS	18.58	.385	16.99	.359	19.53	.168	3	16.92	.350
LEVEL	0.445	.333	0.424	.243	0.465	.100	2	0.440	.440
FLOW	15.02	.453	13.31	.200	14.85	.155	3	13.39	.314

Tab. 2. Achieved minimal RMSE and autocorrelation ρ_e for five real time series.

"Shifted indices" method is the best one for all time series except the **GAS** series. The "sparse indices" method is the worst one in most cases. The classical Holt–Winters method is always less accurate than the method with harmonics.

The optimal number h of full harmonics differs among the individual series. **TEMP** series suffices with $h = 1$ (it is an optimal value) since the monthly average temperature copies a simple sinusoidal curve. On the other hand, for **AIR** series the RMSE gradually goes down as higher values of h are used. This decline stops at the optimal value $h = 5$. It reflects the more complicated seasonal pattern of the series. The remaining three series have $h = 2$ or 3 as their optimum.

See Figures 1, 2 and 3 for the original data, smoothed values and point predictions for **AIR**, **FLOW** and **TEMP** series obtained by the method with trigonometric functions (only the last four periods of data and one future period are displayed). We can see that the method works reasonably – the prediction curves nicely extrapolate the data. Even using $h = 5$ full harmonics for **AIR** series did not lead us to over-parametrization.

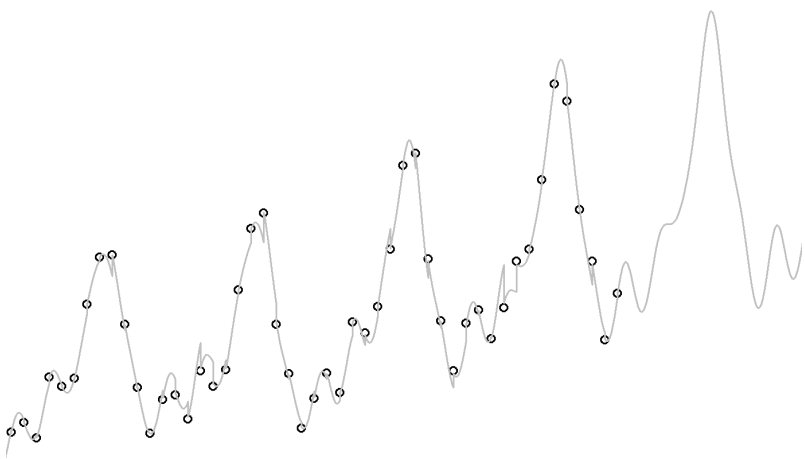

Fig. 1. AIR series: multiplicative Holt–Winters method with 5 full harmonics.

7. SUMMARY

General seasonality modeling concept was suggested in the framework of Holt–Winters method. The multiplicative version is received from the additive one simply by putting logarithms and exponentials in certain formulas. Several particular settings are suggested.

Interpolated seasonal indices can be used routinely to handle general irregular time series. They can also be used to reduce the number of seasonal indices used or to improve the forecasting accuracy by a certain shift of the time axis. Alternatively trigonometric functions (h full harmonics) can be used. This is automatically applicable also for irregular time series and also for regular series it provides a relevant competitor to the classical Holt–Winters method.

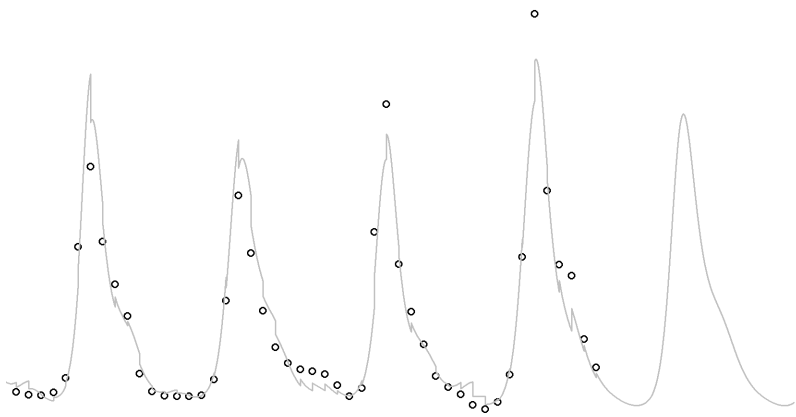

Fig. 2. FLOW series: multiplicative Holt–Winters method with 3 full harmonics.

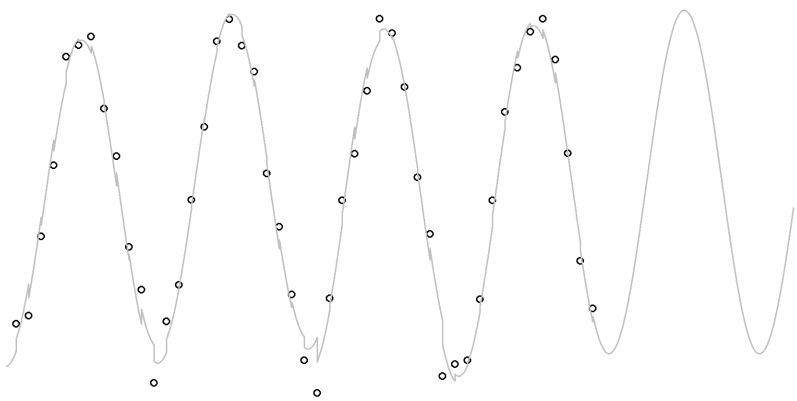

Fig. 3. TEMP series: additive Holt–Winters method with 1 full harmonic.

The suggested methods were successfully tested via simulation study and on real data. In general, seasonal indices outperform trigonometric functions where seasonal jumps, peaks and dips are present. On the other hand, in the case of a smooth seasonal pattern, trigonometric functions (with suitable h) can do better. Sometimes even $h = 1$ can give good results, usually $h = 2$ or 3 is optimal. One should beware of using values $h \gg 5$. Anyway, it usually turns out that the seasonal indices are better choice when the optimal value of h tends to be too large.

In the context of Holt–Winters method the more general and complex model of seasonality does not automatically bring better accuracy even of *in-sample* forecasts (*out-of-sample* forecasts do not surprise us). This is caused by the adaptivity of the seasonal

amplitudes. If we make use of a specific shape of the seasonal pattern (e. g. it is a sinusoidal curve), we can anticipate the next future seasonal component based on the last observed one. This can help us to improve our forecasts.

ACKNOWLEDGEMENT

The work was supported by the grant SVV 261315/2011.

(Received August 7, 2010)

REFERENCES

[1] M. Aldrin and E. Damsleth: Forecasting non-seasonal time series with missing observations. J. Forecasting *8* (1989), 97–116.

[2] C. Chatfield and M. Yar: Holt–Winters forecasting: some practical issues. The Statistician *37* (1988), 129–140.

[3] T. Cipra and T. Hanzák: Exponential smoothing for irregular time series. Kybernetika *44* (2008), 385–399.

[4] T. Cipra, J. Trujillo, and A. Rubio: Holt–Winters method with missing observations. Management Sci. *41* (1995), 174–178.

[5] E. S. Gardner: Exponential smoothing: The state of the art. J. Forecasting *4* (1985), 1–28.

[6] E. S. Gardner: Exponential smoothing: The state of the art – Part II. Internat. J. Forecasting *22* (2006), 637–666.

[7] T. Hanzák: Improved Holt method for irregular time series. In: WDS'08 Proc. Contributed Papers, Part I – Mathematics and Computer Sciences, Matfyzpress, Prague 2008, pp. 62–67.

[8] C. C. Holt: Forecasting seasonals and trends by exponentially weighted moving averages. Internat. J. Forecasting *20* (2004), 5–10.

[9] R. J. Hyndman: Time Series Data Library, www.robhyndman.info/TSDL. Accessed on 26 June 2010.

[10] R. J. Hyndman, A. B. Koehler, R. D. Snyder, and S. Grose: A state space framework for automatic forecasting using exponential smoothing methods. Internat. J. Forecasting *18* (2002), 439–454.

[11] T. Ratinger: Seasonal time series with missing observations. Appl. Math. *41* (1996), 41–55.

[12] J. W. Taylor: Short-term electricity demand forecasting using double seasonal exponential smoothing. J. Oper. Res. Soc. *54* (2003), 799–805.

[13] J. W. Taylor: A comparison of univariate time series methods for forecasting intraday arrivals at a call center. Management Sci. *54* (2008), 253–265.

[14] P. R. Winters: Forecasting sales by exponentially weighted moving averages. Management Sci. *6* (1960), 324–342.

[15] D. J. Wright: Forecasting data published at irregular time intervals using extension of Holt's method. Management Sci. *32* (1986), 499–510.

Tomáš Hanzák, Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics – Charles University, Sokolovská 83, 186 75 Praha 8. Czech Republic.

e-mail: hanzak@karlin.mff.cuni.cz