Some problems of periodic time series
Dissertation progress report

Tomáš Hanzák

seminar
Stochastic modeling in economics and finance
February 28, 2011
Content

1. Dissertation summary
   - My dissertation topic
   - Expected content of the thesis
   - Publications

2. Holt-Winters method with general seasonality
   - Introduction
   - General seasonality
   - Useful special cases
   - Numerical results

3. Supervised theses, teaching

4. References, contacts
Student’s name: Tomáš Hanzák
Doctoral student since: October 2007
Study branch: m-5 Econometrics and operational research
Dissertation title: Some problems of periodic time series
Supervisor: prof. RNDr. Tomáš Cipra, DrSc.
Supervising department: Dept. of Probability and Mathematical Statistics

Annotation:
The doctoral student will familiarize with some time series analysis methods which are modified to be applicable for time series with irregular observations. The stress will be put on modeling periodicity (seasonality) in these time series. The suggested methods will be transposed into the software form.
Looking back...

Topics of my previous presentations here:

1. **2008, May 19**: Improved Holt method for irregular time series
2. **2009, March 16**: Holt-Winters method with general seasonality
3. **2010, March 22**: Exponential smoothing for time series with outliers
4. **2011, February 28**: Summary, Holt-Winters method reminded
Wright (1986) suggested an extension of the Holt method for irregular time series:

\[ S_{t+1} = (1 - \alpha_{t+1}) \cdot [S_t + (t_{t+1} - t_t) \cdot T_t] + \alpha_{t+1} \cdot y_{t+1}, \]
\[ T_{t+1} = (1 - \gamma_{t+1}) \cdot T_t + \gamma_{t+1} \cdot \frac{S_{t+1} - S_t}{t_{t+1} - t_t}, \]
\[ \alpha_{t+1} = \frac{\alpha_t}{\alpha_t + (1 - \alpha)(t_{t+1} - t_t)} \quad \text{and} \quad \gamma_{t+1} = \frac{\gamma_t}{\gamma_t + (1 - \gamma)(t_{t+1} - t_t)}. \]

The division by \( t_{t+1} - t_t \) makes \( T_{t+1} \) very sensitive in the case that \( t_{t+1} - t_t \approx 0 \Rightarrow \text{danger of slope estimate} \ T_{t+1} \text{ destruction.} \)

Suggested modification consists in a modified smoothing coefficient \( \gamma_{t+1} \):

\[ \gamma_{t+1} = \frac{\gamma_t}{\gamma_t + \frac{t_{t+1} - t_t}{t_{t+1} - t_t} (1 - \gamma)(t_{t+1} - t_t)}. \]

The modified method outperformed the original one in a simulation study.
Improved Holt method for irregular time series (2)

Visual illustration:

Forecasts and smoothed values obtained by the original and modified method (both with fixed $\alpha = 0.3$ and $\gamma = 0.1$).
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.
Occurrence of outliers in times series were discussed and conceptual approaches to this problem were listed.

Several particular methods were mentioned and described: Exponential smoothing in $L_1$ norm, discounted M-estimation using IRLS and robust Kalman filter for state space models.

Practically the robust Kalman filter can lead to error truncation: One-step-ahead forecasting error is truncated before entering into the error-correction formulas of the method.

Forecasting errors scale estimator is needed to calibrate the truncation. E.g. GARCH(1, 1) approach can be used.

Error truncation methods were compared with M-estimation in a large simulation study.
Classical exponential smoothing methods (simple exponential smoothing, Holt method, Holt-Winters method) are MSE-optimal for certain ARIMA/SARIMA models.

These models can be taken as a basis for model based approach to construction of extensions of these methods for time series with missing observations.

Aldrin and Damsleth (1989) derived optimal smoothing coefficients for a case of a single gap in observations for simple exponential smoothing (ARIMA(0, 1, 1)) and Holt method (ARIMA(0, 2, 2)).


I derived optimal smoothing coefficients for a general time series with missing observations for a simple exponential smoothing.
1 Introduction

2 Topic overview
   2.1 Periodicity in time series
   2.2 Time series irregularities
   2.3 Overview of approaches

3 Exponential smoothing for irregular time series
   3.1 Methods overview
   3.2 Improved Holt method for irregular time series
   3.3 Holt-Winters method with general seasonality

4 Irregularly observed ARIMA/SARIMA processes

5 Robust methods
   5.1 Methods overview
   5.2 Error truncation

6 Summary
Publications


- T. Cipra, T. Hanzák: Exponential smoothing for time series with outliers. Accepted in Kybernetika.

- T. Hanzák: Holt-Winters method with general seasonality. Accepted in Kybernetika.
Observations, comments...

- My paper about Holt-Winters method was rejected by three journals and then finally accepted on the 4th try ⇒ optimism is needed!

- I had a bad luck and faced the same reviewer (not excited about my paper) twice. Misunderstanding by a reviewer will lead to rejection of the paper ⇒ be explanatory, give references etc. Try to cite potential reviewers and editors in your paper.

- Reply from Appl. Math.: "...clanek nebude publikovan v casopise Appl. Math; stylove totiz nezapada do AM v tom smyslu, ze publikujeme clanky rozclenene do tvrzeni a jejich dukazu, lemmat, definic, atd."

- One Mathematics student from Chile found my previous seminar presentation and asked me for consultancy to his problem (make forecasts in a hourly time series of passenger demand for interurban bus line).

- For me, it is very refreshing to face problems from practice. Unfortunately, ”practice” often does not respect the topic of my dissertation...
Seasonal time series

**Seasonality** = tendency of a certain pattern to repeat over a *fixed* time interval called *season length* or *period*.

Most of time series in practice are seasonal, containing e.g. annual, weekly or daily seasonality:

- **Annual seasonality** is caused by alternation of seasons, annual festivals and holidays or legislation.

- **Weekly seasonality** naturally comes from alternating of working days and weekends.

- **Daily seasonality** is caused by a regular switch between day and night time, regular working or opening hours etc.

The exact period is known in most cases (as above). Sometimes one time series has multiple seasonal components with different periods.

We distinguish basically between *additive and multiplicative seasonality* depending on in which manner the seasonal component is composed with the trend.
Seasonal indices vs. trigonometric functions

Special methods are used to detect seasonality and to model, smooth and forecast seasonal time series. In general, two basic ideas are used in all these situations:

**Seasonal indices**
- ✓ Easy to implement and interpret, any seasonal pattern (even a very non-smooth, with sharp peaks etc.) can be modeled.
- ✗ Large number of parameters needed (danger of overparametrization, sometimes low statistical significance), worse performance in case that the seasonal pattern is of a certain special form (this information is not used), problems with out of grid observations.

**Trigonometric functions**
- ✓ Still sufficiently rich class of seasonal patterns can be created using a lower number of parameters (better statistical significance). Obtained seasonal pattern is continuous by its nature (ideal for irregular time series).
- ✗ More difficult to implement and interpret. Problems with modeling non-smooth seasonal patterns with sharp peaks.
Testing a null hypothesis that the analyzed time series \( \{y_t\} \) of length \( N \) is a gaussian white noise against the alternative that it has a deterministic periodic component (of unknown length).

The test statistic is

\[
g = \frac{\max_{j=1, \ldots, m} I(\omega_j)}{\sum_{j=1}^{m} I(\omega_j)},
\]

where \( I(\omega_j) \) is periodogram of \( \{y_t\} \) evaluated at frequency \( \omega_j = \frac{2\pi j}{N} \) and \( m \) is a lower integer part of \( (N - 1)/2 \).

It considers just the periods of lengths \( N/j \) for \( j = 1, 2, \ldots \). When \( N \) is not an integer multiple the of actual period, the test has a very limited power.

Adding one more observation to the current series \( \Rightarrow N \) changes to \( N + 1 \) \( \Rightarrow \) all the considered \( \omega_j \)'s are changed \( \Rightarrow \) we can get totally different test result.

Usually we suspect certain period lengths (e.g. the integer ones) which we would like \( \{y_t\} \) to be tested against.
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 \( t \).

The **forecast** \( \hat{y}_{t+\tau}(t) \) of \( y_{t+\tau} \), \( \tau > 0 \) from time \( t \) is of the logical form

\[
\hat{y}_{t+\tau}(t) = L_t + \tau \cdot T_t + S_t \oplus \tau,
\]

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 recurrent formulas (in their **error-correction** form)

\[
\begin{align*}
L_{t+1} &= L_t + T_t + \alpha \cdot e_{t+1}, \\
T_{t+1} &= T_t + \alpha \cdot \gamma \cdot e_{t+1}, \\
S_{t+1} &= S_{t-p} + (1 - \alpha) \cdot \delta \cdot e_{t+1},
\end{align*}
\]

where \( e_{t+1} = y_{t+1} - \hat{y}_{t+1}(t) \) is the one-step-ahead forecasting error at time \( t + 1 \) and \( \alpha, \gamma, \delta \in (0,1) \) are the **smoothing constants**.
Holt-Winters method and irregular time series

Holt-Winters method uses **seasonal indices** to deal with seasonality, i.e. it works with \( p \) different indices to form the actual seasonal pattern.

As a consequence, Holt-Winters method gains all the **advantages** and also suffers all the **troubles** connected with using seasonal indices.

One important fact on which the usage of seasonal indices relies, is that each observation **can be assigned to exactly one of \( p \) calendar units** forming the complete season.

This is still possible in a time series with **missing observations**, see Cipra et al. (1995) for such an extension of Holt-Winters method. But the calendar assignment is not possible in **general irregular time series**.

We can use an **interpolation of seasonal indices** to treat the inter-calendar observations. The other possibility is to use **trigonometric functions** which can prevent us also from another disadvantages of seasonal indices.
General seasonality modeling

Seasonality can be generally modeled using $K \geq 1$ different real-valued functions $f_1, f_2, \ldots, f_K$, all defined on the whole real line $\mathbb{R}$. Each $f_k$ is supposed to be periodic with a specific period $p_k \in (0, +\infty)$.

The seasonal pattern is formed as a linear combination of $f_k$:

$$S_t = \sum_{k=1}^{K} A^k \cdot f_k(t),$$

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 to create the smoothed values and forecasts:

$$\hat{y}_t = L_t + S_t.$$

To get a multiplicative seasonality, we put $\hat{y}_t = L_t \cdot \exp[S_t]$.

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.
H-W method with general seasonality modeling

Let \( \{y_{tn}, n \in \mathbb{Z}\} \), \( t_{n+1} > t_n \), be an **irregular seasonal time series** with local linear trend and additive seasonality. We consider its **level** \( L_{tn} \), **slope** \( T_{tn} \) and **seasonal component**

\[
S_{tn}(t) = \sum_{k=1}^{K} A_{tn}^k \cdot f_k(t)
\]

at time \( t_n \). Here \( A_{tn}^k \) are **adaptive amplitudes** at time \( t_n \).

The **forecast** \( \hat{y}_{tn+\tau}(t_n) \) of \( y_{tn+\tau} \), \( \tau > 0 \) from time \( t_n \):

\[
\hat{y}_{tn+\tau}(t_n) = L_{tn} + \tau \cdot T_{tn} + S_{tn}(t_n + \tau)
\]

After a **new observation** \( y_{tn+1} \) becomes available, \( L \), \( T \) and \( A^k \) are updated:

\[
\begin{align*}
L_{tn+1} &= L_{tn} + (t_{n+1} - t_n)T_{tn} + \alpha_{tn+1} e_{tn+1}, \\
T_{tn+1} &= T_{tn} + \alpha_{tn+1} \gamma_{tn+1} e_{tn+1} / (t_{n+1} - t_n), \\
A_{tn+1}^k &= A_{tn}^k + (1 - \alpha_{tn+1}) \delta_{tn+1}^k e_{tn+1} / f_k(t_{n+1}), \quad k = 1, \ldots, K.
\end{align*}
\]

where \( e_{tn+1} = y_{tn+1} - \hat{y}_{tn+\tau}(t_n) \). The **smoothing coefficients** \( \alpha_{tn+1}, \gamma_{tn+1} \) and \( \delta_{tn+1}^k \) will be defined on the next slide.
Smoothing coefficients $\alpha_{tn}$ and $\gamma_{tn}$

Variable smoothing coefficient $\alpha_{tn}$ for **level** is updated in a recurrent way exactly as in Wright (1896) or Cipra et al. (1995):

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

where $\alpha \in (0, 1)$ is the **smoothing constant for level**.

For the smoothing coefficient $\gamma_{tn}$ for **slope**, we will use a modified updating formula **robust to time close observations problem**:

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

where $\gamma \in (0, 1)$ is the **smoothing constant for slope**.
Smoothing coefficients $\delta_{tn}^k$

Variable smoothing coefficients $\delta_{tn}^k$ for the **seasonal amplitudes** are also updated in a recurrent way.

We consider $K$ different **smoothing constants** $\delta^k \in (0, 1)$ belonging to each of the functions $f_k$ and their amplitudes $A_{tn}^k$.

For $k = 1, \ldots, K$ let us denote

$$W_{tn}^k = \sum_{j=0}^{+\infty} \left(1 - \delta^k\right)^{tn-tn-j} \left[f^k(tn-j)\right]^2.$$ 

Obviously $W_{tn}^k$ can be easily updated **recurrently**:

$$W_{tn+1}^k = \left(1 - \delta^k\right)^{tn+1-tn} \cdot W_{tn}^k + \left[f^k(tn+1)\right]^2.$$
Smoothing coefficients $\delta_{tn}^k$ II

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

$$\Delta_{tn+1}^k \equiv f^2_k(t_{n+1}) / W_{tn+1}^k \in [0, 1].$$

Further let us denote

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

$$D_{tn+1} \equiv \sum_{k=1}^K \Delta_{tn+1}^k \geq 0.$$

Finally take the values of the needed smoothing coefficients as

$$\delta_{tn+1}^k \equiv \frac{\Delta_{tn+1}}{D_{tn+1}} \cdot \Delta_{tn+1}^k \in [0, 1], \ k = 1, \ldots, K.$$
Smoothing coefficients $\delta^k_{tn+1}$ as defined above have reasonable properties:

- By $A^k_{tn+1}$ update we move from $\hat{y}_{tn+1}(tn)$ closer to $y_{tn+1}$. Summing the $k$ movements, we come to $S_{tn+1}(tn+1) = S_{tn}(tn+1) + (1 - \alpha_{tn+1})\Delta_{tn+1}e_{tn+1}$. So the total portion of $e_{tn+1}$ absorbed by seasonals is $(1 - \alpha_{tn+1})\Delta_{tn+1} \in [0, 1]$.

- The error $e_{tn+1}$ is absorbed more to $f_k$ with higher $\delta^k$ (i.e. it has really the meaning of a smoothing constant) and with $f_k^2(tn+1)$ larger compared to its recent values.

- If $f_k(tn+1) \to 0$ then (ceteris paribus) $\delta^k_{tn+1}/f_k(tn+1) \to 0$. This means that we do not need to worry about values of $f_k$ near to 0.
Inspiration from DLS linear regression

Let us consider a discounted least squares (DLS) estimation of parameters in the linear regression model $y_t = \sum_{k=1}^{K} A^k f_k(t)$ (without an absolute term) with discount factor $1 - \delta \in (0, 1)$.

Denote $A^k_{t_n}$ the estimates based on data for $t = \ldots, t_{n-2}, t_{n-1}, t_n$. Further denote $\hat{y}_{t_{n+1}}(t_n) = \sum_{k=1}^{K} A^k_{t_n} f_k(t_{n+1})$ the forecast of $y_{t_{n+1}}$ and $e_{t_{n+1}} = y_{t_{n+1}} - \hat{y}_{t_{n+1}}(t_n)$ the forecasting error.

Let $A^k_{t_{n+1}}$ be the estimates of $A^k$ based on the extended data set up to time $t_{n+1}$. Denote $\hat{y}_{t_{n+1}} = \sum_{k=1}^{K} A^k_{t_{n+1}} f_k(t_{n+1})$. We can write

$$A^k_{t_{n+1}} = A^k_{t_n} + \Delta^k_{t_{n+1}} e_{t_{n+1}} / f_k(t_{n+1})$$

for some $\Delta^k_{t_{n+1}} \in \mathbb{R}$.

Given that the left side matrix of the normal system for DLS estimation using data up to $t_{n+1}$ is diagonal (i.e. the regressors are orthogonal), we get

$$\Delta^k_{t_{n+1}} = f^2_k(t_{n+1}) / W^k_{t_{n+1}}$$

with $W^k_{t_{n+1}}$ defined as before. But this is our formula for $\Delta^k_{t_{n+1}}$. 

Tomáš Hanzák

Some problems of periodic time series
Up to now we have considered only \textit{additive} seasonality. To get the \textit{multiplicative} seasonality, one just has to replace the \textbf{prediction formula} with

\[ \hat{y}_{t_n+\tau}(t_n) = [L_{t_n} + \tau \cdot T_{t_n}] \cdot \exp [S_{t_n}(t_n + \tau)] . \]

The recurrent formula for the \textbf{amplitudes update} is now simply changed to

\[ A_{t_{n+1}}^k = A_{t_n}^k + (1 - \alpha_{t_{n+1}}) \delta_{t_{n+1}}^k \left[ \ln y_{t_{n+1}} - \ln \hat{y}_{t_{n+1}}(t_n) \right] / f_k(t_{n+1}) , \quad k = 1, \ldots, K . \]

So just the exponential and logarithm transformations must be placed correctly into two formulas of the method.

Of course we must be sure that both all the observations and forecasts are \textbf{positive}.
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. Generally with higher $K$ we are able to model more precisely even complicated patterns but we must beware of over-fitting.

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

- To set up the **initial values** $L_0$, $T_0$, $\alpha_0$, $\delta_0$, $A_0^k$ and $W_0^k$ before running the recursive computation. We recommend using the general approach of backcasting (**backward forecasting**).
Classical Holt-Winters method

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

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

So $f^k$ are the indicators of individual calendar units and it is $p_k = p$ for all $k = 1, \ldots, K$. Finally take $\delta^k = \delta$ for $k = 1, \ldots, K$.

We get the smoothing coefficients for the seasonal indices in a trivial form:

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

So only one amplitude $A^k$ (belonging to the actual calendar unit of $t$) is updated in one time step.
By taking the functions $f^k$ and the smoothing constants $\delta^k$ exactly the same as in the classical Holt-Winters method and just allowing the analyzed time series $y$ to have missing observations, we come to the same formulas as Cipra et al. (1995).

Only one amplitude $A^k$ (belonging to the actual calendar unit of $t$) is updated in one time step. But now the non-zero smoothing coefficient vary step by step. Its value is determined by the value of the corresponding $W^k$ statistic, containing the information about the time structure of the series when the current calendar unit is concerned.
To treat time series with **general time irregularity**, we must extend the seasonal indices to cover also the inter-calendar observations.

This can be done by a **linear interpolation of the neighboring indices**:

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

- **Classical method**: $K = p$ and $o = 0$. It is an extension of the method for missing observations from Cipra et al. (1995).
- **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.
The seasonal pattern will be composed from several **harmonic curves** of different periods. We will always involve sine and cosine function of the same period. The periods will be taken as \( p, \frac{p}{2}, \frac{p}{3}, \frac{p}{4}, \ldots \).

For example, when \( K = 4 \) (only even values of \( K \) are used), we define

\[
\begin{align*}
f_1(t) &= \sin \frac{2\pi t}{p}, \quad f_2(t) = \cos \frac{2\pi t}{p}, \quad f_3(t) = \sin \frac{4\pi t}{p}, \quad f_4(t) = \cos \frac{4\pi t}{p}.
\end{align*}
\]

The user just have to specify the value of \( h = K/2 \), i.e. **how many full harmonic to use**.

Sometimes even \( h = 1 \) can give good results. The values \( h = 2 \) or 3 are applicable in most cases while with \( h = 4 \) even quite complicated seasonal pattern can be managed (of course the reasonable values for \( h \) are limited from above).

Let us notice that the trigonometric functions \( f_k \) are **centered to 0, linearly independent and approximately orthogonal**.
Further possibilities

Chatfield and Yar (1988) mention the possibility to normalize the seasonal indices in Holt-Winters method to ensure that they always sum up to 0. We can employ such a normalizing in our general seasonality concept.

Often there are multiple seasonal components present in the series. Example: yearly, weekly and daily seasonality in hourly TV audience, electricity demand or SMS traffic series. This can also be represented in our general seasonality concept.

We can even consider non-periodic functions \( f_k \) without any problems. E.g. all Sundays and public holidays can be marked as belonging to one calendar unit in a weekly seasonality.

Exponentially decaying function \( f \) can be used to form an autocorrelated cyclical component in any exponential smoothing method.
Real data examples - data and methods

We have downloaded five regular monthly time series (i.e. containing annual seasonality, \( p = 12 \)):

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

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 \( \alpha, \gamma \) and \( \delta \) 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.
Achieved minimal in-sample RMSE, forecasting errors autocorrelation $\rho_e$ and the optimal value of $h$.

<table>
<thead>
<tr>
<th>Series</th>
<th>Classical H-W</th>
<th>Shifted indices</th>
<th>Sparse indices</th>
<th>Trigonometric</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE $\rho_e$</td>
<td>RMSE $\rho_e$</td>
<td>RMSE $\rho_e$</td>
<td>$h$ RMSE $\rho_e$</td>
</tr>
<tr>
<td>AIR</td>
<td>10.69 .237</td>
<td>10.25 -.124</td>
<td>16.44 -.126</td>
<td>5 10.41 .203</td>
</tr>
<tr>
<td>TEMP</td>
<td>0.740 .180</td>
<td>0.693 .114</td>
<td>0.799 -.181</td>
<td>1 0.713 -.121</td>
</tr>
<tr>
<td>LEVEL</td>
<td>0.445 .333</td>
<td>0.424 .243</td>
<td>0.465 .100</td>
<td>2 0.440 .440</td>
</tr>
</tbody>
</table>
Real data examples - AIR series

**AIR series**: multiplicative Holt-Winters method with 5 full harmonics.
Real data examples - FLOW series

FLOW series: multiplicative Holt-Winters method with 3 full harmonics.
Real data examples - TEMP series

TEMP series: additive Holt-Winters method with 1 full harmonic.
Simulation study - setup

\[ y_t = L_t + S_t + \varepsilon_t, \quad \varepsilon_t \sim iid \ N(0, 1), \]
\[ L_t = L_{t-1} + \mu_t, \quad \mu_t \sim iid \ N(0, 0.1^2), \]
\[ S_t = (1 - \nu) \cdot (S_{t-1} + S_{t-p} - S_{t-p-1}) - \nu \cdot \sum_{j=t-p+1}^{t-1} S_j + \pi_t, \quad \pi_t \sim iid \ N(0, 1) \]

with \( \{\pi_t\}, \{\mu_t\} \) and \( \{\varepsilon_t\} \) mutually independent. The parameter \( \nu \in [0, 1] \) rules the normalization of \( S \) to sum up to 0 and the smoothness of the seasonal pattern (lower \( \nu \) creates a smoother pattern).

For a given \( p \in \{7, 12, 24\} \), we simulate time series of length \( 21p \). The first 10 periods are thrown away. The next 10 periods are used to optimize the smoothing constants \( \alpha \) and \( \delta \) and \( h \) in order to minimize in-sample RMSE of one-step-ahead forecasting errors (we use \( \gamma = 0.05 \) fixed).

The last period is used to calculate out-of-sample RMSE from all the possible \( p(p + 1)/2 \) forecasting errors.

For each \( p \) we use \( \nu = 0.05, 0.1 \) and 0.2 and for each combination of \( p \) and \( \nu \) we simulate 100 time series.
Simulation study - results

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

<table>
<thead>
<tr>
<th></th>
<th>Classical H-W</th>
<th>Shifted indices</th>
<th>Trigonometric</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
</tr>
<tr>
<td>p</td>
<td>v</td>
<td>ranking</td>
<td>ranking</td>
</tr>
<tr>
<td>7</td>
<td>0.05</td>
<td>2.910 1.88</td>
<td>2.786 1.94</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>2.761 1.91</td>
<td>2.624 1.85</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>2.195 2.01</td>
<td>2.185 2.17</td>
</tr>
<tr>
<td>12</td>
<td>0.05</td>
<td>3.626 2.11</td>
<td>3.347 1.77</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>3.551 2.04</td>
<td>3.180 1.95</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>2.291 1.86</td>
<td>2.243 1.80</td>
</tr>
<tr>
<td>24</td>
<td>0.05</td>
<td>7.294 2.21</td>
<td>3.246 1.45</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>4.023 2.12</td>
<td>2.870 1.46</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>2.189 1.48</td>
<td>2.222 1.76</td>
</tr>
</tbody>
</table>
Supervised bachelor theses


Aleh Masaila: Regression trees

The development of computer technologies provides us many new methods for obtaining useful information from available data. One of such methods are so called regression trees. The aim of the diploma thesis will be to describe this method as an alternative to other techniques used for identification and estimation of the influence of the independent variables on the dependent variable (e.g. linear or logistic regression, discriminant analysis etc.). The focus will be put on identification of the weaknesses and strengths of different approaches and their comparison. A demonstration of the practical application of regression trees on a selected data set will be part of the work.

Ivana Meňhartová: Methods of dynamical analysis of portfolio composition

Diplomantka popíše problém dynamické analýzy složení portfolia na základě jeho pozorovaných výnosů. Představí existující metody řešení tohoto problému (rolující regrese a Kalmanův filtr), které otestuje na reálných, simulovaných či uměle vytvořených datech. Výsledky bude podrobně analyzovat a interpretovat a poukáže na slabé a silné stránky jednotlivých metod v různých situacích. Cílem bude navrhnut vhodné modifikace existujících metod tak, aby bylo dosaženo co nejlepších výsledků z hlediska porovnání odhadovaného a skutečného vývoje složení portfolia v čase.
Teaching

  One group of exercises for informatics students.

- Summer term 2010/11: **Time series - exercises** (NSTP007).
  One group of exercises (optional for attendants of the lecture).

- Summer term 2010/11: **Credit risk in banking** (NFAP042).
  Lecture shared with Miloš Kopa.
References


Tomáš Hanzák

mobile: 604 799 879
e-mail: tomas.hanzak@post.cz
web: www.thanzak.sweb.cz

Department of Probability and Mathematical Statistics
Faculty of Mathematics and Physics
Charles University in Prague
Sokolovská 83, 186 75 Praha 8.
e-mail: hanzak@karlin.mff.cuni.cz
web: www.karlin.mff.cuni.cz/~kpms

MEDIARESEARCH, a.s.
Českobratrská 1, 130 00 Praha 3.
mobile: 725 535 535
e-mail: tomas.hanzak@mediaresearch.cz
web: www.mediaresearch.cz