Charles University in Prague

Faculty of Mathematics and Physics

DOCTORAL THESIS

Mgr. Tomáš Hanzák

Methods for periodic and irregular time series

Department of Probability and Mathematical Statistics

Supervisor of the doctoral thesis:	Prof. RNDr. Tomáš Cipra, DrSc.
Study programme:	Mathematics
Specialization:	Econometrics and Operational Research

Prague 2014

First of all I must express my sincere gratitude to my supervisor Prof. Tomáš Cipra for his extraordinarily kind, patient and tireless guidance and support throughout my whole doctoral study.

My thanks go also to the rest of the pedagogues, to the reviewers of my published papers and to all who have somehow contributed to my thesis.

Further, I would like to thank to my parents and my whole family for their unconditional love, support and trust in me.

I also would like to mention my good friends including several of my colleagues from MEDIARESEARCH company whose presence means a lot to me as well.

Last but not least, I would like to dedicate my thesis to my dear girlfriend, who does not have a great understanding of mathematics but always had a great understanding for me.

Acknowledgement

The work is a part of the research project MSM0021620839.

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

I declare that I carried out this doctoral thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In Prague, 31st March 2014

Mgr. Tomáš Hanzák

Název práce: Metody pro periodické a nepravidelné časové řady

Autor: Mgr. Tomáš Hanzák

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí disertační práce: Prof. RNDr. Tomáš Cipra, DrSc.

Abstrakt: Disertační práce se primárně zabývá modifikacemi metod typu exponenciální vyrovnávání pro jednorozměrné časové řady s periodicitou a/nebo určitými typy nepravidelností. Je navržena modifikovaná Holtova metoda pro nepravidelné časové řady robustní vůči problému "časově blízkých" pozorování. Obecný koncept modelování sezónnosti je zaveden do Holtovy-Wintersovy metody včetně lineární interpolace sezónních indexů a použití goniometrických funkcí jako speciálních případů (obě metody jsou použitelné pro nepravidelná pozorování). Je zkoumán DLS odhad regrese s lineárním trendem a sezónními indexy a metoda je porovnána s aditivní Holtovou-Wintersovou metodou. Autokorelovaný člen je navržen jako další složka dekompozice časové řady. Navržené metody jsou porovnávány s klasickými na reálných datech a/nebo prostřednictvím simulačních studií.

Klíčová slova: Diskontované nejmenší čtverce, exponenciální vyrovnávání, Holtova-Wintersova metoda, nepravidelná pozorování, periodicita časových řad

Title: Methods for periodic and irregular time series

Author: Mgr. Tomáš Hanzák

Department: Department of Probability and Mathematical Statistics

Supervisor: Prof. RNDr. Tomáš Cipra, DrSc.

Abstract: The thesis primarily deals with modifications of exponential smoothing type methods for univariate time series with periodicity and/or certain types of irregularities. A modified Holt method for irregular times series robust to the problem of "time-close" observations is suggested. The general concept of seasonality modeling is introduced into Holt-Winters method including a linear interpolation of seasonal indices and usage of trigonometric functions as special cases (the both methods are applicable for irregular observations). The DLS estimation of linear trend with seasonal dummies is investigated and compared with the additive Holt-Winters method. An autocorrelated term is introduced as an additional component in the time series decomposition. The suggested methods are compared with the classical ones using real data examples and/or simulation studies.

Keywords: Discounted Least Squares, Exponential smoothing, Holt-Winters method, Irregular observations, Time series periodicity

Contents

List of Notations

$\mathbb{R}$, $\mathbb{N}$, $\mathbb{Z}$	Sets of real, natural and whole numbers
$x \equiv y$	x equals y, x is defined as y
$x \approx y$	x approximately equals y
$x \gg y$	x is much greater than y
$x \to y$	x tends to y
$t \mod p$	t modulo p
x^+	Positive part of x, i.e. $x^+ \equiv \max(0, x)$
$\lfloor x \rfloor$	Integer part of real number x
$\boldsymbol{A}, \boldsymbol{v}$	Matrix, column vector
$\boldsymbol{A}', \boldsymbol{v}'$	Matrix and vector transposition
$\mathrm{Diag}(\cdot)$	Diagonal matrix with given elements
$\boldsymbol{A}^{-1}$	Matrix inversion
$\det \boldsymbol{A}$	Determinant of square matrix $\boldsymbol{A}$
$\mathsf{N}(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2
$\mathsf{N}_n(\mathbf{0}, \boldsymbol{R})$	n-dimensional normal distribution with mean $\mathbf{0}$ and covariance $\boldsymbol{R}$
u_θ	θ-quantile of $\mathsf{N}(0, 1)$
U	Uniform distribution
δ_x	Dirac measure at point x
t	Time
$\varepsilon_t \sim iid\ (\mathcal{L})$	ε_t are independent and identically distributed (with distribution $\mathcal{L}$)
$\varepsilon_t, \mu_t, \pi_t, \eta_t, \theta_t$	iid disturbances (innovations) at time t
t_n	Time of the nth observation of irregular time series
q	Average time spacing of time series
y_t	Observation of time series y at time t
Δy_t, $\Delta^2 y_t$	First and second difference of time series y at time t
B	Backshift time series operator, $By_t = y_{t-1}$

$\hat{y}_t$	Smoothed value of time series y at time t
$\hat{y}_{t+\tau}(t)$	Forecast of value $y_{t+\tau}$, $\tau > 0$, from time t
τ	Forecasting horizon
p	Period length of time series
$t \oplus \tau$	Latest time not later than t which is congruent with $t + \tau \bmod p$
e_t	One-step-ahead forecasting error at time t
v_t	Forecast variance factor at time t
L_t, T_t	Level and slope of time series at time t
S_t	Seasonal component of time series at time t
C_t	Autocorrelated component of time series at time t
$S_t^{[i]}$	ith order smoothing statistics at time t
α, γ, δ	Smoothing constants
α_H, γ_H	Smoothing constants (for level and slope) of Holt method
α_{HW}, γ_{HW}, δ_{HW}	Smoothing constants of Holt-Winters method
α_{t_n}, γ_{t_n}	Smoothing coefficients for level and slope at time t_n
$\delta_{t_n}, \delta_{t_n}^k$	Smoothing coefficients for seasonal component at time t_n
φ	Damping constant
β	Discount factor
ν	Parameter controlling smoothness of seasonal pattern
f_k	Real valued p_k-periodic function on $\mathbb{R}$
K	Number of periodic functions f_k in seasonal component
$\boldsymbol{F}_n$	Regression design matrix up to time n
$\boldsymbol{D}_n$	Diagonal discounting matrix up to time n
A^k, A_t^k	Amplitude of f_k in seasonal component of time series (at time t)
W_t^k	Exponential weighted average of squared values of function f_k
$f_k \circ f_l$	Discounted scalar product of functions f_k and f_l
$\overline{f_k^2}$	Average squared value of f_k over available observation times
o	Time origin of interpolated seasonal indices
h	Number of full harmonics in Holt-Winters method

List of Abbreviations

AR	**A**uto**R**egressive
ARIMA	**A**uto**R**egressive **I**ntegrated **M**oving **A**verage
DES	**D**ouble **E**xponential **S**moothing
DLS	**D**iscounted **L**east **S**quares
DMITS	**D**ecomposition **M**ethods for **I**rregular **T**imes **S**eries
EWMA	**E**xponential **W**eighted **M**oving **A**verage
H-W	**H**olt-**W**inters method
IRLS	**I**teratively **R**eweighted **L**east **S**quares
MA	**M**oving **A**verage
MSE	**M**ean **S**quare **E**rror
RMSE	**R**oot **M**ean **S**quare **E**rror
SARIMA	**S**easonal **A**uto**R**egressive **I**ntegrated **M**oving **A**verage
SES	**S**imple **E**xponential **S**moothing
WLS	**W**eighted **L**east **S**quares

Preface

This doctoral thesis summarizes my research as a Ph.D. student at Charles University in Prague, Department of Probability and Mathematical Statistics, in the period 2007-2014.

I would like to thank to my supervisor Prof. Tomáš Cipra, inter alia, for his valuable contribution to my research and thus to this thesis. However, all the potential errors and deficiencies of it must be attributed solely to me personally.

The topic of my dissertation focusing on irregular time series follows the topic of my diploma thesis *Decomposition methods for time series with irregular observations* written under the supervision of Prof. Cipra in the period 2005-2007 at the same university, see Hanzák (2007). The elements of this diploma thesis were then published in Hanzák and Cipra (2008). In the dissertation this is mentioned in Sections 2.2 and 2.3 of the existing methods survey.

A substantial part of the original material in this dissertation (covered by Chapters 3 and 4) has been published already, see Hanzák (2008) and Hanzák (2012). The content of Chapters 5 and 6 has not been published so far.

One should start to read the thesis ideally with Introduction and Chapters 1 and 2 containing the topic overview and the *state of the art*. The later Chapters 3, 4, 5 and 6, containing the author's own contribution to the topic, are mostly thematically independent and can be read without any problems in an arbitrary order.

The software implementation of majority of the methods suggested was done by the author. The application is called DMITS (**D**ecomposition **M**ethods for **I**rregular **T**ime **S**eries) and it is an extended version of the application originally developed as a part of author's diploma thesis, see Chapter 7. The application is included as an electronic attachment to the thesis.

The electronic attachment further includes the real and simulated time series data used in Sections 3.5, 3.6, 4.5 and 6.5, all the figures of the thesis as source PNG files and all the tables in XLS file.

In Prague, 31st March 2014

Mgr. Tomáš Hanzák

Introduction

An amount of data available for statistical analysis is growing over time both in terms of data volumes and number of different kinds of data sources. This process is nowadays driven mainly by a rapid development of information technology and its penetration to real lives of millions of people.

It is definitely unnecessary to prove and illustrate the obvious fact that *time series* is a type of statistical data largely occurring in almost every scientific discipline or field of human activity. Simply said, almost every data can be ordered and aggregated to form a time series. And since we live in an era of constant changes, it is usually useful to do so.

The information technologies not just bring new data to us but it also offer a database and computational power to handle and analyze them. There is a remarkable development in *time series analysis* as a field of econometrics and in time series software available as well. It shows that a proper understanding (both factual and statistical) and analytical skills of the data owners/users are the natural most limiting factors.

Real time series sometimes exhibit various types of ”irregularities”: missing observations, observations collected not regularly over time or outlying observations. This dissertation primarily focuses on time series smoothing and forecasting methods modified for usage with such ”irregular” time series. We restrict ourselves to univariate real-valued time series with emphasis on periodic time series. The methods considered can all be called as variants of *exponential smoothing*, i.e. they are of a decomposition, adaptive and recursive nature.

Some of the classical exponential smoothing methods like those of Holt and Winters (see Winters (1960) or Holt (2004)) were already extended for application in the context of irregular time series, see Wright (1986), Aldrin and Damsleth (1989), Cipra et al. (1995), Ratinger (1996), Cipra (2006), Hanzák and Cipra (2008), Croux et al. (2008) and Gelper et al. (2010). However, some combinations of irregularity types and methods to be applied were still not covered or the extended methods available suffer from certain defects in their performance.

The goal of this dissertation is to fill some of these gaps and suggest additional methods usable for these cases. All the suggested methods are illustrated and compared with the existing ones using real data examples and/or simulation studies.

The text of the thesis is organized as follows:

Chapter 1 more closely describes various types of time series irregularities and difficulties they bring to us. Special attention is paid to time series with observations irregularly spanned over time, with missing observations as its special case. Periodicity in time series is then discussed in light of these irregularities.

Chapter 2 is a *state of the art* - an overview of existing methods of exponential smoothing (or similar) modified for irregular time series. This includes simple exponential smoothing (SES) and Holt method by Wright (1986), Holt-Winters methods by Cipra et al. (1995), double exponential smoothing (DES) from Cipra (2006) and exponential smoothing of order m from Hanzák and Cipra (2008). Methods derived from assuming the fully observed series to follow a certain ARIMA or SARIMA model are reviewed, see Aldrin and Damsleth (1989), Ratinger (1996) and Hanzák and Cipra (2008). State space modeling and Kalman filter is mentioned as a general powerful tool for filtering, smoothing and forecasting in time series. Some robust exponential smoothing methods are mentioned as well.

The following chapters 3, 4, 5, 6 and 7 contain the author's own contributions to the dissertation topic.

An improvement of Holt method for irregular observation times by Wright (1986) that is robust to a problem of time-close observations (two subsequent observations with time distance much shorter than in average) is provided in Chapter 3. If this situation occurs in time series when using the original Wright's formulas, one can obtain seriously wrong results. A real data example and a simulation study is provided to compare the performance of the original and modified method. The content of this chapter was published in Hanzák (2008).

Chapter 4 suggests a generalization of Holt-Winters 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 applicable for time series with irregular observations). A simulation study and real data examples compare the suggested methods with the classical one. The content of this chapter was published in Hanzák (2012).

In Chapter 5 we investigate a possibility of *Discounted Least Squares* (DLS) estimation of linear regression with linear trend and seasonal dummies. It is shown that this leads approximately to a certain special case of Holt-Winters method with additive seasonality. Analytical formulas expressing the respective

Holt-Winters smoothing constants in the terms of the discount factor used and period length of the series are derived, visualized and commented.

Chapter 6 considers an *autocorrelated component* added to the decomposition schemes of classical exponential smoothing type methods. It is a semi-systematical component responsible for autocorrelated variation around the series trend. The extended version of simple exponential smoothing (also applicable for irregular time series) is introduced in detail, the case of a general method is then provided as well. A numerical example of the application of the extended method is provided.

Author's software application DMITS (**D**ecomposition **M**ethods for **I**rregular **T**ime **S**eries), implementing most of the existing and newly suggested methods, is described in Chapter 7. The application itself is contained in the electronic attachment to the thesis.

Chapter 1

Problem properties of time series

In this chapter the topic of the dissertation is outlined from the point of view of the various time series irregularities and the challenges for time series methods caused by them. Particular existing methods are reviewed in Chapter 2.

In Section 1.1 the problem of time series observations spread irregularly over time is introduced. Section 1.2 focuses on periodicity in time series. Section 1.3 mentions time series with outliers.

1.1 Time series with irregular observations

There are many more or less different ways how a time series can be formed:

- A series of equally time spaced measurements of a variable which is defined in continuous time. E.g. a daily time series of total TV audience (% of people watching TV) in 4+ target group at 8 p.m.

- A series of aggregated volumes of a certain activity within regularly spaced equally long time intervals. E.g. a weekly time series of counts of advertising spots broadcasted on a particular TV channel.

- A result of a different type of aggregation which is not a simple "summing". E.g. a daily time series of number of viewers of a particular TV channel (so called *daily reach* of a TV channel).

- A series of observed numerical characteristic of a periodically occurring phenomenon. E.g. a daily time series of number of viewers of main news program of a particular TV channel.

However, in all the cases above the time series observations can be naturally attributed to a regular (equally spaced) time grid, so these are all classical *regular* time series. Since this happens in most cases in practise, the vast majority of methods of time series analysis are designed for regular time series only.

Nevertheless, there are situations where we have a time series formed from values observed at irregularly spaced times, i.e. the observation times can form a general increasing sequence, not just an arithmetic one. We are talking about a time series with *irregular observations* or briefly about an *irregular time series*. Regular time series can be viewed as a special case of irregular time series.

Sometimes the irregularity is just very slight, for example caused by a different length of calendar months. Or the irregularity occurs very few times in the analyzed time series. In such cases we can simply neglect the time irregularity of the series and we won't get results significantly affected.

Time series with *missing observations* (certain amount of observations of a regular time series are not available) form a kind of intermediate stage between regular and irregular time series. A necessary condition to talk about time series with missing observations is that all observation time steps (difference of consecutive observation times) are multiples of a basic time step. So there exist irregular time series that cannot be interpreted as series with missing observations. Analogously there are models and methods that are applicable to the case of missing observations but not to the general case of irregular time series.

If the missing observations occur sparsely in a times series, we can try to fill in the missing observations somehow (by expert estimates or using some sort of interpolation) and then apply a method suitable for regular time series.

In time series with missing observations, the time steps are always integer (or multiples of a basic time step). Especially the time step cannot be arbitrarily small. The same holds also for the forecasting horizons we are interested in. When having a general irregular time series, the time step is an arbitrary positive real number which can be arbitrarily small. This can cause problems to some methods, see Chapter 3.

Seasonality in time series is often modeled by so called *seasonal indices*, i.e. a set of seasonal dummies. This becomes non-trivial when there is a general time irregularity in the time series. See Sections 1.2 and 2.1 and Chapter 4 for details.

Various examples of time series with irregular observations can be found in Wright (1986):

- **Change in reporting frequency**. When a statistical office or other institution increases the frequency of publication of a variable, e.g. from annual to quarterly, the resulted overall time series has irregular observation times.

- **Irregular measurements**. Sometimes due to objective reasons it is impossible or expensive to guarantee regularly spaced measurements times. Or the observations come from more different sources with different observation frequency.

- **Observation times driven by the series itself**. E.g. cumulative counts of occurrences of a disease in a given area (observation time = occurrence date). Or a time series of men world record times for 1 mile (observation time = date of establishing a new world record), see Section 3.5.

An example of the second type comes from establishment/calibration survey of Czech TV audience measurement project 2007-2012. This survey was organized in 10 "monthly" waves per year. There was no wave in December due to Christmas (households not available for interviewing) and there was one summer wave taking place in July and August (due to summer holidays). Results of a particular question (e.g. penetration of home internet connection or digital terrestrial TV reception) from individual waves formed an irregular times series.

Qing et al. (2010) employed, among others, the methods from Wright (1986) to construct short-term traffic forecasts using data irregularly collected from GPS devices placed in a sample of 480 taxis in Hong Kong.

To end up the section, we will look at examples of time series where it is not so clear whether we should treat them as regular or as irregular time series:

- A times series of daily closing prices of a stock title on a stock exchange where only working days are trading days. The resulting time series looks like a typical series with missing observations (weekends and holidays). But these missing values have never existed, not that they were just not measured.

- A time series of consecutive waiting times for a certain event to occur again and again, e.g. a bet placed through an online betting system by a customer.

1.2 Periodicity in time series

Lot of time series from practise are of seasonal (periodic) character. Seasonal component is characterized by a fixed *period length* after which a similar pattern repeats itself over time. This is in contrast to a cyclical component (e.g. in a macroeconomic variable) where the variation does not have a fixed period length.

Annual seasonality is primarily caused by alternation of seasons, annual festivals and holidays or legislation. Weekly seasonality naturally comes from alternating of weekdays and weekends. Daily seasonality is caused by a regular switch between day and night time, regular working or opening hours etc. We distinguish between *additive* and *multiplicative* seasonality depending on in which manner the seasonal component is composed with the trend.

The exact period is known in most cases. Sometimes one time series has multiple seasonal components with different periods. E.g. Taylor (2003) considered a time series of half-hour electricity consumptions which has at least a daily and weekly periodicity. Similarly TV audience rating has daily (prime time vs. off time), weekly (weekdays vs. weekends) and yearly (changing of weather and daylight) periodicity. Moreover, these different seasonal components often interact with each other: e.g. daily TV audience pattern is different on weekdays and weekends. Thus forecasting a time series with such a complex seasonality demands always a lot of effort to find a method with reasonably low number of seasonality parameters (components) but capturing the seasonal pattern in its whole complexity.

Two basis approaches are used to deal with seasonality, i.e. to help us to detect (test), model, smooth and forecast seasonality or adjust for it. These are *seasonal indices* and *trigonometric functions*. They both have their advantages and disadvantages.

Seasonal indices are easy to implement and interpret. Any seasonal pattern (even a very non-smooth, with sharp falls, peaks etc.) can be modeled in this way. However, large number of seasonal indices is needed when the period length is large (danger of over-parametrization, problems with statistical significance).

Trigonometric functions are more difficult to implement and interpret. They can still usually form a sufficiently rich class of seasonal patterns using a lower number of parameters (better statistical significance). However, they typically fail to model very non-smooth seasonal patterns.

Irregularity of observation time grid brings specific problems to seasonality modeling. Intermediate case of time series with missing observations can still be handled in a quite similar way as regular time series. In regular time series, each observation could be assigned to exactly one of the finite number of seasons (January, February etc. in the case of annual seasonality). This doesn't work anymore for irregular time series. Here an observation can occur whenever in the seasonal period, i.e. even somewhere "between" the two seasons (if any "seasons" are defined at all). Here it seems that trigonometric functions have the advantage of a continuous seasonal pattern they form.

In fact the time series analyzed does not even need to be irregular - it suffices that the series period length is not a multiple of its basic observation time step. As an example, consider a weekly time series with annual seasonality. Since one year does not have a whole number of weeks, this situation cannot be solved by seasonal indices in a straightforward way.

1.3 Outliers in time series

Outliers in time series are an example of defect or irregularity in time series data. By an "outlier" in time series it is possible to call several slightly different things:

- An absolutely outlying observation due to bug in data processing. These are maybe rather academical examples. In practice, such bugs should be primarily avoided, not managed by the smoothing and forecasting methods applied to data.
- A correct observation which is however quite far away from the remaining data bulk or trend. Result of an occasional extreme event in the measured reality (catastrophe, political intervention, panic at the market etc.).
- An observation that is by chance lying little bit "outside". The underlying distribution can have fat tail(s).

Most of time series methods like regression fitting or exponential smoothing are linear in the sense that the smoothed values and forecasts are conditional expectations estimated by linear functions of the available observations. So such methods are sensitive to presence of outliers in the analyzed time series which can have negative impact on smoothing and forecasting results.

In case of outliers' presence in the analyzed time series, one can proceed in several ways:

- To identify and remove these outlying observations. Treat the remaining data as a time series with missing observations.
- To identify outliers and to replace them by values interpolated from the remaining "inlying" values. Then apply the classical method on the new time series.
- To make the classical method somehow robust against outliers.

Chapter 2

Survey of existing methods

In this chapter some methods for coping with irregular time series, existing prior to those suggested in my dissertation, are briefly described. This serves as a "state of the art" overview and also forms a basis for describing the newly suggested methods since they are often modifications or extensions of these already existing methods.

Most of the methods described here belongs to the broad family of exponential smoothing. This in the simplest case means that they put weights decreasing exponentially into past to the observations of the series. This makes the methods adaptive, i.e. they are able to track a changing level, slope and/or seasonal pattern of the series. Taking the weights as exponential function of time also enables the method to be formulated in terms of relatively simple recursive formulas. This makes them easy to implement and understand. An explanatory introduction to the idea of exponential smoothing can be found in Hanzák (2007). A comprehensive overview of exponential smoothing methods for regular time series is provided by Gardner (1985, 2006).

Some natural extensions of these exponential smoothing methods to the case of irregular time series or time series with missing observations have been presented in past. As well as the original methods, these extensions are often based on *ad hoc* idea of exponential weighting which proves its relevance by good empirical performance of the methods.

Alternatively, stochastic models were specified for which the original methods are optimal and following this line, methods for time series with missing observations were derived. This represents a *model-based* approach.

Which approach is the "right" one is not clear and it depends on several circumstances and preferences of the analyst. Model-based methods have an advantage of a deeper theoretical background, useful e.g. for calculation of prediction errors and prediction intervals construction. On the other hand, *ad hoc* extensions of the original methods have maybe a better chance to be accepted by practitioners - especially when they need to implement the method by themselves. When the forecasting accuracy is concerned, usually the both approaches seem to be comparable.

The methods briefly described in this chapter are:

- Section 2.1: Simple exponential smoothing (SES) and Holt method for irregular time series by Wright (1986).
- Section 2.1: Holt-Winters method for time series with missing observations by Cipra et al. (1995).
- Section 2.1: Holt method with exponential or damped linear trend for irregular time series by Cipra (2006).
- Section 2.2: Double exponential smoothing (DES) for irregular time series by Cipra (2006).
- Section 2.2: Exponential smoothing of order m for irregular time series by Hanzák and Cipra (2008).
- Section 2.3: SES, Holt method and DES by Aldrin and Damsleth (1989) for the case of a single gap in observations, assuming the series to be driven by ARIMA(0, 1, 1) and ARIMA(0, 2, 2) models.
- Section 2.3: Holt-Winters method for the case of a single gap in observations by Ratinger (1996), assuming the series to be driven by appropriate SARIMA model.
- Section 2.3: SES for irregular time series by Hanzák and Cipra (2008). ARIMA(0, 1, 1) model is assumed to derive the optimal smoothing coefficient for each time step.
- Section 2.4: State space models accompanied by *Kalman filter* as a powerful general tool for smoothing and forecasting time series.
- Section 2.5: Robust exponential smoothing methods: Exponential smoothing in L_1 norm and M-estimation by Cipra (1992), general approach of error truncation, see Gelper et al. (2010) or Cipra and Hanzák (2011).

In the following sections, we will consider either a regular time series $\{y_t, t \in \mathbb{Z}\}$ or an irregular time series $\{y_{t_n}, n \in \mathbb{Z}\}$ with observation times $t_{n+1} > t_n,\ n \in \mathbb{Z}$. By $\hat{y}_t$ we will denote the smoothed value of the series y at time t and by $\hat{y}_{t+\tau}(t)$ the forecast of $y_{t+\tau}$ constructed at time t ($\tau > 0$ is the forecasting horizon). Finally let us denote $e_{t+1} = y_{t+1} - \hat{y}_{t+1}(t)$ the one-step-ahead forecasting error at time $t+1$ (from time t). Analogous notation will be used also for irregular time series.

2.1 Simple exponential smoothing, Holt and Holt-Winters method

Classical methods nowadays referred to as *exponential smoothing* were developed in the late 50s to predict future sale volumes of goods for the optimal management of their production and storage. The idea of using the concept of exponential weighting to estimate not just the level of the time series but its trend and seasonal components as well, was published first by American Charles C. Holt in 1957 in his memorandum to the Office of Naval Research, see Holt (2004). The suggested methods together with the application context can be found in Winters (1960).

Simple exponential smoothing (SES) is the simplest exponential smoothing method and it is in some sense a basic building block for the more complex methods. It is designed for time series with locally constant trend. The level of the series at time t is denoted as L_t.

Classical SES for regular time series consists of following smoothing, forecasting and recursive updating equations:

$$\hat{y}_t = \hat{y}_{t+\tau}(t) = L_t, \quad \tau > 0\,, \tag{2.1}$$

$$L_{t+1} = (1-\alpha)L_t + \alpha y_{t+1} = L_t + \alpha e_{t+1}\,, \tag{2.2}$$

where $\alpha \in (0,1)$ is a fixed smoothing constant. The second form of (2.2) is called an *error correction* form. An initial value of L must be chosen to start the recursion, usually as an average (possibly weighted) of a couple of initial observations of the series. Value of $\alpha \in (0,1)$ is usually tuned to optimize certain in-sample forecasting accuracy criterion like Root Mean Square Error (the same holds analogously for the next two method described, Holt and Holt-Winters method).

Wright (1986) suggested an extension of this method for the case of time series observed at irregular time intervals:

$$\hat{y}_{t_n} = \hat{y}_{t_n+\tau}(t_n) = L_{t_n}, \quad \tau > 0\,, \tag{2.3}$$

$$L_{t_{n+1}} = (1-\alpha_{t_{n+1}})L_{t_n} + \alpha_{t_{n+1}} y_{t_{n+1}} = L_{t_n} + \alpha_{t_{n+1}} e_{t_{n+1}}\,, \tag{2.4}$$

$$\alpha_{t_{n+1}} = \frac{\alpha_{t_n}}{\alpha_{t_n} + (1-\alpha)^{t_{n+1}-t_n}}\,. \tag{2.5}$$

Here α_{t_n} is a time varying smoothing coefficient reflecting the structure of observation times and still $\alpha \in (0,1)$. Initial value of α_{t_n} is taken as fixed point of formula (2.5) with $t_{n+1} - t_n = q$, where $q > 0$ is the average time spacing of time series y:

$$\alpha_{t_0} = 1 - (1 - \alpha)^q \,. \tag{2.6}$$

Holt method is an exponentially smoothing method designed for time series with locally linear trend proposed by Holt (2004) and Winters (1960). It considers its level L_t and slope T_t at time t. There are two smoothing constants, $\alpha \in (0,1)$ for level and $\gamma \in (0,1)$ for slope.

Classical Holt method for regular time series consists of the following formulas:

$$\hat{y}_t = L_t \,, \tag{2.7}$$

$$\hat{y}_{t+\tau}(t) = L_t + \tau T_t \,, \tag{2.8}$$

$$L_{t+1} = (1-\alpha)(L_t + T_t) + \alpha y_{t+1} = L_t + T_t + \alpha e_{t+1} \,, \tag{2.9}$$

$$T_{t+1} = (1-\gamma)T_t + \gamma(L_{t+1} - L_t) = T_t + \gamma\alpha e_{t+1} \,. \tag{2.10}$$

Initial values of L and T are chosen e.g. as regression line coefficients fitted through a couple of initial observations of the series (DLS estimation can be used to put more weight to the observations at the very beginning of the series).

Wright (1986) suggested an extension of this method for irregular time series:

$$\hat{y}_{t_n} = L_{t_n} \,, \tag{2.11}$$

$$\hat{y}_{t_n+\tau}(t_n) = L_{t_n} + \tau T_{t_n} \,, \tag{2.12}$$

$$L_{t_{n+1}} = (1 - \alpha_{t_{n+1}}) \left[L_{t_n} + (t_{n+1} - t_n) T_{t_n} \right] + \alpha_{t_{n+1}} y_{t_{n+1}} \,, \tag{2.13}$$

$$T_{t_{n+1}} = (1 - \gamma_{t_{n+1}}) T_{t_n} + \gamma_{t_{n+1}} \frac{L_{t_{n+1}} - L_{t_n}}{t_{n+1} - t_n} \,. \tag{2.14}$$

Time varying smoothing coefficients α_{t_n} and γ_{t_n} are initialized and updated in the same way as in the case of Wright's SES:

$$\alpha_{t_{n+1}} = \frac{\alpha_{t_n}}{\alpha_{t_n} + (1-\alpha)^{t_{n+1}-t_n}} \quad \text{and} \quad \gamma_{t_{n+1}} = \frac{\gamma_{t_n}}{\gamma_{t_n} + (1-\gamma)^{t_{n+1}-t_n}} \,. \tag{2.15}$$

Initial values of L and T are constructed in the same way as in the regular case.

Holt method with exponential trend is a modification of the original Holt method with (locally) linear trend, see e.g. Gardner (1985). For many economic time series this is a more natural choice: in short term horizon rather the relative differences are stable than the absolute ones.

Cipra (2006) followed the idea of Wright (1986) and formulated Holt method with exponential trend for irregular time series:

$$\hat{y}_{t_n+\tau}(t_n) = L_{t_n} T_{t_n}^{\tau} \,, \tag{2.16}$$

$$L_{t_{n+1}} = (1-\alpha_{t_{n+1}}) L_{t_n} T_{t_n}^{t_{n+1}-t_n} + \alpha_{t_{n+1}} y_{t_{n+1}} \,, \tag{2.17}$$

$$T_{t_{n+1}} = (1-\gamma_{t_{n+1}}) T_{t_n} + \gamma_{t_{n+1}} \left(\frac{L_{t_{n+1}}}{L_{t_n}} \right)^{1/(t_{n+1}-t_n)} \,. \tag{2.18}$$

Time varying smoothing coefficients α_{t_n} and γ_{t_n} remain the same, see (2.15). The method for regular time series is easily obtained as a special case for regular observation times t_n. Similar (but not equivalent) results can be obtained via applying the classical Holt method to the series of logarithms and then apply exponential transformation to the results.

Holt method with damped linear trend is another modification of the original Holt method invited by Gardner (1985). It follows the idea that the current trend slope is usually damped in future. The modification anticipates this and thus provides more conservative forecasts in middle and long term. Slope decay is supposed to be exponential with *damping parameter* $\varphi \in (0,1)$. The version of this method for irregular time series was presented by Cipra (2006):

$$\hat{y}_{t_n+\tau}(t_n) = L_{t_n} + g(\tau) T_{t_n} \,, \tag{2.19}$$

$$L_{t_{n+1}} = (1-\alpha_{t_{n+1}}) \left[L_{t_n} + g(t_{n+1}-t_n) T_{t_n} \right] + \alpha_{t_{n+1}} y_{t_{n+1}} \,, \tag{2.20}$$

$$T_{t_{n+1}} = \varphi^{t_{n+1}-t_n} \left[(1-\gamma_{t_{n+1}}) T_{t_n} + \gamma_{t_{n+1}} \frac{L_{t_{n+1}} - L_{t_n}}{g(t_{n+1}-t_n)} \right] \,, \tag{2.21}$$

where we have denoted $g(x) \equiv \varphi \frac{1-\varphi^x}{1-\varphi}$ for $x > 0$. It is $g(k) = \sum_{i=1}^{k} \varphi^i$ for $k \in \mathbb{N}$, i.e. $g(x)$ expresses the accumulation of slope being damped over time. The coefficients α_{t_n} and γ_{t_n} again remain the same as in (2.15).

It would not be a problem to formulate Holt method with *damped exponential* trend as well. However, Holt method with damped linear trend together with the logarithm-exponential transformation will do a very similar job.

Holt-Winters method is Holt method enriched with *seasonal indices* S_t to reflect a seasonal pattern of period p in the series, being either additive or multiplicative. Let us start with the additive case for regular series. It is

$$\hat{y}_{t+\tau}(t) = L_t + \tau T_t + S_{t \oplus \tau} \,, \tag{2.22}$$

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

$$L_{t+1} = (1-\alpha)(L_t + T_t) + \alpha(y_{t+1} - S_{t+1-p}), \tag{2.23}$$

$$T_{t+1} = (1-\gamma)T_t + \gamma(L_{t+1} - L_t), \tag{2.24}$$

$$S_{t+1} = (1-\delta)S_{t+1-p} + \delta(y_{t+1} - L_{t+1}), \tag{2.25}$$

where $\alpha, \gamma, \delta \in (0,1)$ are smoothing constants for level, slope and seasonal indices. Equations (2.23)-(2.25) are again often rewritten to their equivalent *error-correction* form, see e.g. Gardner (1985, 2006):

$$L_{t+1} = L_t + T_t + \alpha e_{t+1}, \tag{2.26}$$

$$T_{t+1} = T_t + \gamma\alpha e_{t+1}, \tag{2.27}$$

$$S_{t+1} = S_{t+1-p} + \delta(1-\alpha)e_{t+1}. \tag{2.28}$$

Version with multiplicative seasonality differs in these formulas:

$$\hat{y}_{t+\tau}(t) = (L_t + \tau T_t)S_{t\oplus\tau}, \tag{2.29}$$

$$L_{t+1} = (1-\alpha)(L_t + T_t) + \alpha y_{t+1}/S_{t+1-p}, \tag{2.30}$$

$$S_{t+1} = (1-\delta)S_{t+1-p} + \delta y_{t+1}/L_{t+1}. \tag{2.31}$$

An extension of the method for time series with missing observations was provided by Cipra et al. (1995). Its additive variant consists of

$$L_{t_{n+1}} = (1-\alpha_{t_{n+1}})\left[L_{t_n} + (t_{n+1} - t_n)T_{t_n}\right] + \alpha_{t_{n+1}}(y_{t_{n+1}} - S_{t^*_{n+1}}), \tag{2.32}$$

$$T_{t_{n+1}} = (1-\gamma_{t_{n+1}})T_{t_n} + \gamma_{t_{n+1}}\frac{L_{t_{n+1}} - L_{t_n}}{t_{n+1} - t_n}, \tag{2.33}$$

$$S_{t_{n+1}} = (1-\delta_{t_{n+1}})S_{t^*_{n+1}} + \delta_{t_{n+1}}(y_{t_{n+1}} - L_{t_{n+1}}), \tag{2.34}$$

where t^*_{n+1} is the largest value among $t_n, t_{n-1}, \ldots$ such that t^*_{n+1} belongs to the same season as t_{n+1}. Time varying smoothing coefficients α_{t_n} and γ_{t_n} remain the same, see (2.15). Seasonal smoothing coefficient $\delta_{t_{n+1}}$ is updated as

$$\delta_{t_{n+1}} = \frac{\delta_{t^*_{n+1}}}{\delta_{t^*_{n+1}} + (1-\delta)^{(t_{n+1} - t^*_{n+1})/p}}. \tag{2.35}$$

The forecasts are of the form

$$\hat{y}_{t_n+\tau}(t_n) = L_{t_n} + \tau T_{t_n} + S_{(t_n+\tau)^*} \,, \tag{2.36}$$

where $(t_n+\tau)^*$ is the largest value among $t_n, t_{n-1}, \ldots$ such that $(t_n+\tau)^*$ belongs to the same season as $t_n + \tau$. The multiplicative case differs in these formulas:

$$\begin{aligned}
\hat{y}_{t_n+\tau}(t_n) &= (L_{t_n} + \tau T_{t_n}) S_{(t_n+\tau)^*} \,, && (2.37)\\
L_{t_{n+1}} &= (1-\alpha_{t_{n+1}})\,[L_{t_n} + (t_{n+1}-t_n)T_{t_n}] + \alpha_{t_{n+1}} y_{t_{n+1}}/S_{t^*_{n+1}} \,, && (2.38)\\
S_{t_{n+1}} &= (1-\delta_{t_{n+1}}) S_{t^*_{n+1}} + \delta_{t_{n+1}} y_{t_{n+1}}/L_{t_{n+1}} \,. && (2.39)
\end{aligned}$$

The method works for time series with missing observations but not for general irregular time series due to the need to assign each observation to exactly one of p seasons. Holt-Winters method applicable for irregular time series is suggested in Chapter 4.

2.2 Double exponential smoothing and smoothing of order m

Double exponential smoothing (DES, also known as *Brown method*) is an alternative to Holt method: it fits a regression line through the time series observations using *Discounted Least Squares* (DLS) estimation with discount factor $\beta = 1-\alpha \in (0,1)$ ($\alpha \in (0,1)$ is a smoothing constant). When we consider a time series with infinite history observed, the two regression parameters (intercept and slope) can be expressed as a fixed linear combinations of so called first and second *smoothing statistics* of the series:

$$S_t^{[1]} = \alpha \sum_{j=0}^{\infty} y_{t-j}\beta^j \quad \text{and} \quad S_t^{[2]} = \alpha \sum_{j=0}^{\infty} S_{t-j}^{[1]}\beta^j \,. \tag{2.40}$$

And what is important, the two smoothing statistics can be calculated recursively:

$$S_{t+1}^{[1]} = (1-\alpha)S_t^{[1]} + \alpha y_{t+1} \quad \text{and} \quad S_{t+1}^{[2]} = (1-\alpha)S_t^{[2]} + \alpha S_{t+1}^{[1]} \,. \tag{2.41}$$

DES with smoothing constant $\alpha \in (0,1)$ is in fact equivalent to Holt method with smoothing constants α_H and γ_H given by

$$\alpha_H = \alpha(2-\alpha) \quad \text{and} \quad \gamma_H = \frac{\alpha}{2-\alpha} \,, \tag{2.42}$$

see Mikulka (2008) for the derivation of this relation. So the DES is less flexible than Holt method (α_H and γ_H are tight together) but on other hand, the one-dimensional selection of α is easier and "safer" than the two-dimensional selection of α_H and γ_H. See Mikulka (2008) for further results regarding the relation of the two methods.

Double exponential smoothing for irregular time series was provided by Cipra (2006). Smoothing statistics $S^{[1]}$ and $S^{[2]}$ are defined for irregular time series in the same manner as SES was extended for irregular time series by Wright (1986). Besides smoothing coefficient α_{t_n}, one has to recalculate recursively additional two statistics.

The method can be derived in this way. Consider the values of $S^{[1]}$ and $S^{[2]}$ at the time of the last available observation of the series as values of two different linear operators applied to the series. The goal is to find a linear fit to the time series such that this fit gives the same values of $S^{[1]}$ and $S^{[2]}$ as the series itself. This is a system of two linear equations for two parameters (intercept and slope), an analogue to the system of two normal equations in DLS regression.

For regular time grid, matching the values of first two smoothing statistics is equivalent to DLS regression with $\beta = 1 - \alpha$. For irregular time grid this equivalence does not hold exactly so there are two different ways how to generalize Brown method for irregular times series. However, these two methods are still pretty similar when both computational complexity and empirical performance are concerned. A third possibility (equivalent to none of the two mentioned already) would be to use Holt method by (Wright, 1986) with smoothing constants α_H and γ_H tight according to (2.42).

Exponential smoothing or order m for irregular time series was described in Hanzák and Cipra (2008). Is is a straightforward extension of the above described approach to the case of polynomial fit of order m to the series. Again, this can be done by using DLS regression or fitting the values of the first $m + 1$ smoothing statistics $S^{[p]}$ ($S^{[p+1]}$ is obtained by smoothing $S^{[p]}$). Both variants result in a system of $m + 1$ equations for $m + 1$ parameters of the fitted polynomial. Again, for regular time series the two variants are equivalent but for irregular time series in general not.

The computational demand of these methods is naturally growing with higher orders m. However for $m = 0, 1, 2$ the formulas are still quite simple and easy to implement. The case $m = 0$ corresponds to the Wright's SES (here the both approaches coincide). The case $m = 1$ leads to the DES by Cipra (2006) or similar DLS method. The case $m = 2$ which is the last one with some practical importance is a *triple* exponential smoothing for time series with locally quadratic trend.

2.3 Irregularly observed ARIMA and SARIMA processes

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. This section reviews several methods obtained using this approach.

Simple exponential smoothing with a single gap in observations was studied by Aldrin and Damsleth (1989). When

$$y_{t+1} = L_t + e_{t+1}\,, \tag{2.43}$$

$$L_{t+1} = (1-\alpha)L_t + \alpha\, y_{t+1} \tag{2.44}$$

and the one-step-ahead forecasting errors $\{e_t, t \in \mathbb{Z}\}$ form a white noise (with variance $\sigma^2 > 0$) then

$$\Delta y_t = e_t + (\alpha - 1)e_{t-1}\,, \tag{2.45}$$

i.e. series $\{y_t, t \in \mathbb{Z}\}$ follows an invertible ARIMA(0, 1, 1) process parameterized by $\alpha \in (0, 1)$, the optimal smoothing constant for the regular time series.

Aldrin and Damsleth (1989) assumed this ARIMA(0, 1, 1) model to hold and for such a case derived the value of optimal smoothing coefficient to be used instead of α after a gap in observations. They came to

$$\alpha_{t_{n+1}} = \frac{\alpha^2(t_{n+1} - t_n - 1) + \alpha}{\alpha^2(t_{n+1} - t_n - 1) + 1}\,, \tag{2.46}$$

where t_{n+1} is the current observation time and t_n is the previous one. It is $\alpha_{t_{n+1}} = \alpha$ for $t_{n+1} = t_n + 1$ and $\alpha_{t_{n+1}} \to 1$ for $t_{n+1} - t_n \to \infty$.

According to Aldrin and Damsleth (1989), their method works better than simple exponential smoothing without adjusting the smoothing constant and comparably to the SES by Wright (1986).

ARIMA(0, 1, 1) model with missing observations was studied by Hanzák and Cipra (2008). It is an extension of Aldrin and Damsleth (1989), taking a time structure of the whole series history into account. Let L_t be still the "true" level of the series and $\widetilde{L}_t$ the level received by applying the smoothing method to the series with missing observations. Let us denote

$$v_t \equiv \frac{\mathsf{var}(L_t - \widetilde{L}_t)}{\sigma^2} \,. \tag{2.47}$$

Then the optimal smoothing coefficient is

$$\alpha_{t_{n+1}} = \frac{v_{t_n} + \alpha^2(t_{n+1} - t_n - 1) + \alpha}{v_{t_n} + \alpha^2(t_{n+1} - t_n - 1) + 1} \tag{2.48}$$

and the variance factor v_t is updated by

$$v_{t_{n+1}} = (1 - \alpha_{t_{n+1}})^2[v_{t_n} + \alpha^2(t_{n+1} - t_n - 1)] + (\alpha - \alpha_{t_{n+1}})^2 \,. \tag{2.49}$$

When $v_{t_n} = 0$, (2.48) reduces itself to (2.46). From (2.48) it can be seen that $\alpha_{t_{n+1}} \geq \alpha$ and $\alpha_{t_{n+1}}$ is an increasing function of arguments v_{t_n}, α and $t_{n+1} - t_n$ which is consistent with our intuitive view. For $\alpha \to 1$, $t_{n+1} - t_n \to \infty$ or $v_{t_n} \to \infty$ we have $\alpha_{t_{n+1}} \to 1$. If $v_{t_n} = 0$ and $t_{n+1} - t_n = 1$ then $\alpha_{t_{n+1}} = \alpha$.

Variance of the forecasting error $e_{t_n+\tau}(t_n) = y_{t_n+\tau} - \widetilde{L}_{t_n}$ is

$$\mathrm{var}\left[e_{t_n+\tau}(t_n)\right] = \sigma^2 \left[v_{t_n} + \alpha^2(\tau - 1) + 1\right] \,. \tag{2.50}$$

Although this method has been explicitly derived only for time series with missing observations, it can be used in practice for general irregular time series (never mind that $t_{n+1} - t_n$ is not integer). The method's forecasting accuracy is comparable to that of Wright (1986), see Hanzák and Cipra (2008).

Holt method with a single gap in observations was studied by Aldrin and Damsleth (1989). It is based on the fact that Holt method with smoothing constants α and γ is optimal for a time series driven by ARIMA(0, 2, 2) model

$$\Delta^2 y_t = e_t + (\alpha + \gamma\alpha - 2)e_{t-1} + (1 - \alpha)e_{t-2} \,. \tag{2.51}$$

After a single gap in observations, the optimal smoothing coefficients are

$$\alpha_{t_{n+1}} = \frac{\alpha^2(\Delta t - 1)G + \alpha}{\alpha^2(\Delta t - 1)G + 1}\,, \tag{2.52}$$

$$\gamma_{t_{n+1}} = \frac{\gamma\Delta t\,[1 + \alpha(\Delta t - 1)(1 + \gamma\Delta t/2)]}{\alpha(\Delta t - 1)G + 1}\,, \tag{2.53}$$

where

$$G = 1 + \gamma\Delta t\left[1 + \frac{\gamma(2\Delta t - 1)}{6}\right] \quad \text{and} \quad \Delta t = t_{n+1} - t_n\,. \tag{2.54}$$

If $\Delta t = 1$ then $\alpha_{t_{n+1}} = \alpha$ and $\gamma_{t_{n+1}} = \gamma$. If $\gamma = 0$ then $\gamma_{t_{n+1}} = 0$ and (2.52) reduces to (2.46). As $\Delta t \to \infty$, $\alpha_{t_{n+1}} \to 1$ and $\gamma_{t_{n+1}} \to 1.5$. Limit of $\gamma_{t_{n+1}}$ higher than 1 looks strange at the first moment but it is logically interpreted by the authors[1].

The method performs comparably well in relation to that of Wright (1986). DES as a special case of Holt method is also covered by formulas (2.52) and (2.53). These results could be extended for the general case of missing observations analogously to Hanzák and Cipra (2008) for the case of SES. However, the formulas would be rather complicated and we do not expect the resulted method to offer any practical advantage over the methods already mentioned.

Holt-Winters method with a single gap in observations was studied by Ratinger (1996). It is an extension of the approach undertaken by Aldrin and Damsleth (1989) for SES and Holt method. It can be shown that a series for which the Holt-Winters method with additive seasonality of period p with smoothing constants α, γ and δ is optimal is driven by certain SARIMA model. More specifically, $(1 - B)(1 - B^p)y_t$ follows certain MA(p+1) model which parameters can be expressed in terms of α, γ and δ.

Optimal smoothing coefficients $\alpha_{t_{n+1}}$, $\gamma_{t_{n+1}}$ and $\delta_{t_{n+1}}$ to be used at time t_{n+1} after a single gap in observations were derived by Ratinger (1996). The formulas for $\alpha_{t_{n+1}}$ and $\gamma_{t_{n+1}}$ are rather complicated, involving also the forecasting horizon for which the next forecast should be optimal. Formula for $\delta_{t_{n+1}}$ is much simpler and it is similar to (2.46):

$$\delta_{t_{n+1}} = \frac{\delta^2(1 - \alpha)\lfloor\frac{\Delta t - 1}{p}\rfloor + \delta}{\delta^2(1 - \alpha)\lfloor\frac{\Delta t - 1}{p}\rfloor + 1}\,, \tag{2.55}$$

where $\lfloor x \rfloor$ denotes the integer part of x. Again $\gamma_{t_{n+1}} \to 1.5$ as $\Delta t \to \infty$.

[1] The mean slope between t_n and t_{n+1} could be expected to lie inside the interval determined by the two slopes at t_n and t_{n+1}.

This method is an alternative to Holt-Winters method for time series with missing observations suggested by Cipra et al. (1995). Only the case of additive seasonality is described by Ratinger (1996) but one can try using the same smoothing coefficients also together with a multiplicative seasonality. Both methods are still not possible to apply on a general irregular time series due to usage of seasonal indices. Holt-Winters method applicable for irregular time series is suggested in Chapter 4.

2.4 State space models and Kalman filter

Many of the methods described already can be approached as special cases of linear state space model. The components of the series form the state vector, its evolution over time is specified (including random innovations) and the observations of the time series are functions of the state vectors and random observation errors.

Let us consider a simple discrete time linear state space model of the form[1]

$$\boldsymbol{S}_{t+1} = \boldsymbol{A}_t \boldsymbol{S}_t + \boldsymbol{a}_{t+1}\,, \quad \boldsymbol{a}_t \sim iid\ \mathsf{N}_n(\boldsymbol{0}, \boldsymbol{R}_t)\,, \tag{2.56}$$

$$y_t = \boldsymbol{h}_t' \boldsymbol{S}_t + \varepsilon_t\,, \quad \varepsilon_t \sim iid\ \mathsf{N}(0, r_t^2)\,, \tag{2.57}$$

where $\boldsymbol{S}_t$ is the n-dimensional state vector of the system (do not confuse with the seasonal component or smoothing statistics), $\{y_t\}$ is the one-dimensional observation process, $\{\varepsilon_t\}$ is the observation noise, $\{\boldsymbol{a}_t\}$ is the n-dimensional innovation noise process, $\boldsymbol{A}_t$ is $n \times n$ deterministic matrix, $\boldsymbol{h}_t$ is n-dimensional deterministic vector and $\boldsymbol{R}_t$ and $r_t^2 > 0$ describe the variance-covariance structure of $\boldsymbol{a}_t$ and ε_t, see e. g. Abraham and Ledolter (1983) or Cipra (2008), p. 462–465.

The simple exponential smoothing method relates to the following state space model (so called *random walk plus noise*):

$$y_t = L_t + \varepsilon_t\,, \quad \varepsilon_t \sim iid\ \mathsf{N}(0, \sigma_\varepsilon^2)\,, \tag{2.58}$$

$$L_t = L_{t-1} + \eta_t\,, \quad \eta_t \sim iid\ \mathsf{N}(0, \sigma_\eta^2) \tag{2.59}$$

with ε_t and η_t mutually independent. Here $\boldsymbol{S}_t = L_t$, $\boldsymbol{A}_t = 1$, $\boldsymbol{h}_t = 1$, $\boldsymbol{R}_t = \sigma_\eta^2$ and $r_t^2 = \sigma_\varepsilon^2$.

[1]Matrices and vectors are printed in bold. Vectors are always column vectors.

Holt method relates to the local linear trend model of the form

$$
\begin{aligned}
y_t &= L_t + \varepsilon_t\,, \quad \varepsilon_t \sim iid\, \mathsf{N}(0, \sigma_\varepsilon^2)\,, &(2.60)\\
L_t &= L_{t-1} + T_{t-1} + \eta_t\,, \quad \eta_t \sim iid\, \mathsf{N}(0, \sigma_\eta^2)\,, &(2.61)\\
T_t &= T_{t-1} + \theta_t\,, \quad \theta_t \sim iid\, \mathsf{N}(0, \sigma_\theta^2)\,. &(2.62)
\end{aligned}
$$

Innovation terms η_t and θ_t are mutually independent and also independent of the noise term ε_t. It is $\boldsymbol{S}_t = \begin{pmatrix} L_t \\ T_t \end{pmatrix}$, $\boldsymbol{A}_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, $\boldsymbol{h}_t = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\boldsymbol{R}_t = \begin{pmatrix} \sigma_\eta^2 & 0 \\ 0 & \sigma_\theta^2 \end{pmatrix}$ and $r_t^2 = \sigma_\varepsilon^2$.

The updating equations referred to as Kalman filter are

$$
\begin{aligned}
\hat{\boldsymbol{S}}_{t+1|t} &= \boldsymbol{A}_t \hat{\boldsymbol{S}}_{t|t}\,, &(2.63)\\
\boldsymbol{P}_{t+1|t} &= \boldsymbol{A}_t \boldsymbol{P}_{t|t} \boldsymbol{A}_t' + \boldsymbol{R}_t\,, &(2.64)\\
\hat{\boldsymbol{S}}_{t+1|t+1} &= \hat{\boldsymbol{S}}_{t+1|t} + \boldsymbol{k}_{t+1}\left(y_{t+1} - \boldsymbol{h}_{t+1}' \hat{\boldsymbol{S}}_{t+1|t}\right)\,, &(2.65)\\
\boldsymbol{P}_{t+1|t+1} &= \boldsymbol{P}_{t+1|t} - \boldsymbol{k}_{t+1} \boldsymbol{h}_{t+1}' \boldsymbol{P}_{t+1|t}\,, &(2.66)\\
\boldsymbol{k}_{t+1} &= \frac{\boldsymbol{P}_{t+1|t} \boldsymbol{h}_{t+1}}{\boldsymbol{h}_{t+1}' \boldsymbol{P}_{t+1|t} \boldsymbol{h}_{t+1} + r_t^2}\,, &(2.67)
\end{aligned}
$$

where $\hat{\boldsymbol{S}}_{r|s}$ is an estimate of $\boldsymbol{S}_r$ based on the observations of y up to time s, $\boldsymbol{P}_{r|s}$ is its estimation error covariance matrix and $\boldsymbol{k}_{t+1}$ is called *gain* vector. Since the one-step-ahead prediction of y_{t+1} from time t is naturally $\hat{y}_{t+1|t} = \boldsymbol{h}_{t+1}' \hat{\boldsymbol{S}}_{t+1|t}$, see (2.57), one can rewrite (2.65) to the form

$$
\hat{\boldsymbol{S}}_{t+1|t+1} = \hat{\boldsymbol{S}}_{t+1|t} + \boldsymbol{k}_{t+1}(y_{t+1} - \hat{y}_{t+1|t}) = \hat{\boldsymbol{S}}_{t+1|t} + \boldsymbol{k}_{t+1} e_{t+1}\,, \qquad (2.68)
$$

where $e_{t+1} = y_{t+1} - \hat{y}_{t+1|t}$ are the corresponding prediction errors. The estimates and predictions delivered by this Kalman filter are optimal in the MSE sense.

The models for SES and Holt methods presented above are so called *multi source of error* since they contain several independent innovation and observation error terms. However, the error-correction formulas of exponential smoothing methods contain just one error term, the one-step-ahead prediction error, whose different multiples account for the stochastic dynamics of each model equation (both state and observation). See e.g. Hyndman et al. (2002) or De Livera et al. (2011) for details regarding single source of error state space models.

Missing observations in time series can be handled in several way when using state space models:

- When observation y_t is missing, define $\boldsymbol{h}_t = 0$ and $y_t = 0$. So the artificial observation y_t does not provide any information about the state vector and so it has no impact.

- When observation y_t is missing, set $r_2 = \infty$ (very large positive number) and y_t equal to its forecast constructed at the previous time step. Since the observation y_t is supposed to be contaminated by an infinite observation error, its value has no impact.

- For irregular time series, try to reformulate the state evolution matrix $\boldsymbol{A}_t$ and the innovations covariance matrix $\boldsymbol{R}_t$ to reflect the length of the time step.

2.5 Robust exponential smoothing

Problem of outliers presence in time series was already mentioned in Section 1.3. All the exponential smoothing methods in their classical variants are either based on Discounted Least Squares (DLS), presented *ad hoc* directly in the form of Exponentially Weighted Moving Averages (EWMA) or equivalently based on a recursive application of convex linear combinations of old and new information. So these methods are naturally sensitive to presence of outliers in the analyzed time series.

In Section 1.3 we have also outlined some basic approaches how to deal with outliers. Several robustifications of classical exponential smoothing methods have been suggested in literature.

Exponential smoothing in L_1 norm uses discounted medians (simple exponential smoothing) or discounted regression medians (double exponential smoothing) instead of means, see Cipra (1992). I.e., the minimized criterion is not a discounted sum of *squared* deviations (L_2 norm) but a discounted sum of *absolute* deviations (L_1 norm). This makes the method robust to outliers, paying a little loss of efficiency for normally distributed data and a higher computational effort.

Exponential smoothing using M-estimation instead of Least Squares was suggested by Cipra (1992) and further developed by Croux et al. (2008).

Both the simple and double exponential smoothing follow the same idea of discounted M-estimation provided by *Iteratively Reweighted Least Squares* (IRLS) algorithm. This is a favorite iterative estimation technique transferring a general minimization problem (e.g. M-estimation) into the Weighted Least Squares (WLS) problem using the weights depending on the parameter's estimates from the previous iteration.

To keep the exponential smoothing methods recursive and computationally simple (with no need for multiple iterations), the IRLS algorithm is followed only approximately. In each iteration, instead of recalculation of all the weights, a new observation is included and its weight is assigned based on the trend fitted in the previous time step. The remaining weights are not recalculated, just discounted in time.

The symmetric quasi-convex *loss function* ρ and the discount factor $\beta \in (0, 1)$ must be specified. The weight w_{t+1} assigned to observation y_{t+1} (classical non-robust method is obtained by taking $w_t \equiv 1$) is calculated as

$$w_{t+1} = \frac{s_t \cdot \psi \left[\frac{y_{t+1} - \hat{y}_{t+1}(t)}{s_t} \right]}{y_{t+1} - \hat{y}_{t+1}(t)}, \tag{2.69}$$

where $\psi(x) = \rho'(x)$ and s_t is a scale estimate for the one-step-ahead forecasting error $e_{t+1} = y_{t+1} - \hat{y}_{t+1}(t)$.

Truncation of the prediction errors is used by Gelper et al. (2010) and Cipra and Hanzák (2011). This approach is generally applicable and easily interpreted. Huber ψ-function

$$\psi(x) = \begin{cases} x & \text{if } |x| \leq u_{1-\theta/2} \\ \text{sign}(x) \cdot u_{1-\theta/2} & \text{otherwise} \end{cases} \tag{2.70}$$

($u_{1-\theta/2}$ is the normal $(1 - \theta/2)$-quantile) is used to truncate the normalized prediction errors. By means of the value $\theta \in (0, 1)$ the level of robustness of the method can be tuned.

Various robustifications of Kalman filter represent a possible way how to derive robust smoothing and forecasting methods, see Ruckdeschel et al. (2014) for similar approach to Cipra and Hanzák (2011).

Chapter 3

Holt method and problem of time close observations

3.1 Introduction

At the end of 50's, C. C. Holt and P. R. Winters proposed an *ad hoc* method for smoothing and forecasting time series with locally linear trend, see 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[1].

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. For example, the version with damped linear trend by Gardner and McKenzie (1985) is said to give better results in longer forecasting horizons. All the above mentioned methods were already briefly described in Section 2.1.

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 formulas to this situation (see also Section 2.1) and has tested his method on several real time series. The formulas of the extended version of Holt method by Wright (1986) are in detail reminded in Section 3.2.

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

[1]Sometimes both the methods are referred as *Holt-Winters method* plus telling whether its non-seasonal or seasonal version is considered.

consequent observations have much shorter time distance compared to the average time spacing in the series.

Using Wright's extension of Holt method, this phenomenon can easily cause a rapid shift of slope estimate and consequently a bias in forecasts for a quite long time period. Detailed discussion of this impact is provided in Section 3.3.

In Section 3.4 we suggest a reasonable modification of Wright's formulas to overcome these difficulties. This modification is not less intuitive or easy to understand than the original method. In author's opinion it has even a better interpretation. Moreover, as far as the resulting formulas are concerned, the modification means just adding one term in an updating formula for slope smoothing coefficient.

Section 3.5 contains one real data example (one mile run men world record time series) illustrating the impact of the time-close observations on the original and modified Holt method (DES and DLS linear trend are examined as well).

In Section 3.6 a simulation study is provided to further evaluate the impact of the suggested modification empirically. Several settings of time series generating scheme are employed to find out how the improvement depends on the configuration of time-close observations in the data. Section 3.7 brings the conclusions of the chapter.

3.2 Holt method by Wright

Both the classical Holt method and the version for irregular time series were described briefly in Section 2.1. Remind that we work with the level L_t and the slope T_t of the series at time t. For a regular time series $\{y_t, t \in \mathbb{Z}\}$ the updating formulas are

$$L_{t+1} = (1-\alpha)(L_t + T_t) + \alpha y_{t+1}\,, \tag{3.1}$$

$$T_{t+1} = (1-\gamma)T_t + \gamma(L_{t+1} - L_t)\,, \tag{3.2}$$

where $\alpha \in (0,1)$ is a smoothing constant for level and $\gamma \in (0,1)$ is a smoothing constant for slope. Formula (2.10) is a recurrent version of the exponential weighting

$$T_t = \gamma \sum_{j=0}^{\infty} (1-\gamma)^j \tilde{T}_{t-j}\,, \tag{3.3}$$

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

Extension of this method for the case of time series observed at irregular time intervals was suggested by Wright (1986). He followed the idea of exponential weighting and generalized formulas (2.9) and (2.10) into

$$L_{t_{n+1}} = (1-\alpha_{t_{n+1}})\left[L_{t_n} + (t_{n+1}-t_n)T_{t_n}\right] + \alpha_{t_{n+1}} y_{t_{n+1}}\,, \tag{3.4}$$

$$T_{t_{n+1}} = (1-\gamma_{t_{n+1}})T_{t_n} + \gamma_{t_{n+1}} \frac{L_{t_{n+1}} - L_{t_n}}{t_{n+1}-t_n}\,, \tag{3.5}$$

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

$$\alpha_{t_{n+1}} = \frac{\alpha_{t_n}}{\alpha_{t_n} + (1-\alpha)^{t_{n+1}-t_n}} \quad \text{and} \quad \gamma_{t_{n+1}} = \frac{\gamma_{t_n}}{\gamma_{t_n} + (1-\gamma)^{t_{n+1}-t_n}}\,. \tag{3.6}$$

Again, formula (3.5) is a recurrent version of the underlying exponential weighting

$$T_{t_n} = \gamma_{t_n} \sum_{j=0}^{\infty} (1-\gamma)^{t_n - t_{n-j}} \tilde{T}_{t_{n-j}}\,, \tag{3.7}$$

where

$$\gamma_{t_n} = \left[\sum\nolimits_{j=0}^{\infty} (1-\gamma)^{t_n - t_{n-j}}\right]^{-1} \tag{3.8}$$

plays the role of a normalizing factor and

$$\tilde{T}_{t_k} = \frac{L_{t_k} - L_{t_{k-1}}}{t_k - t_{k-1}} \tag{3.9}$$

is again the one-step slope from time t_{k-1} to t_k.

3.3 Problem of time-close observations

As was already stressed in Section 3.1, 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 $y_{t_{n+1}} - y_{t_n}$ or $L_{t_{n+1}} - L_{t_n}$ can be significantly non-zero. Wright (1986) used several examples of irregular time series, one of them was the time series of men world record times in one mile run, being a good example of a time series with time-close observations, see Section 3.5.

It is good to note that what really matters is not the *absolute* distance between the two observations. When the time axis is re-scaled so that all the time steps are multiplied by a certain scaling factor $\omega > 0$, the method just adjusts its smoothing constants from α and γ to $1-(1-\alpha)^{1/\omega}$ and $1-(1-\gamma)^{1/\omega}$ and the method provides exactly the same results. So the time steps can be arbitrarily short without a negative impact on the method. What matters is the length of the given time step *relative* to the others.

Let us look in detail how the presence of time-close observations affects results obtained by Wright's version of Holt method. From formulas (2.12)-(2.14) one can easily derive the following *error-correction* form of (2.14):

$$T_{t_{n+1}} = T_{t_n} + \frac{\gamma_{t_{n+1}} \alpha_{t_{n+1}}}{t_{n+1} - t_n} e_{t_{n+1}} \,, \tag{3.10}$$

where

$$e_{t_{n+1}} = y_{t_{n+1}} - \hat{y}_{t_{n+1}}(t_n) = y_{t_{n+1}} - [L_{t_n} + (t_{n+1} - t_n) T_{t_n}] \tag{3.11}$$

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

$$\alpha_{t_{n+1}} \to \frac{\alpha_{t_n}}{\alpha_{t_n} + 1} > 0 \quad \text{and} \quad \gamma_{t_{n+1}} \to \frac{\gamma_{t_n}}{\gamma_{t_n} + 1} > 0 \,, \tag{3.12}$$

see (2.15), and so

$$\frac{\gamma_{t_{n+1}} \alpha_{t_{n+1}}}{t_{n+1} - t_n} \to \infty \,. \tag{3.13}$$

Together with the fact that the forecasting error $e_{t_{n+1}}$ in (3.10) is not in any sense restricted to tend to 0 when $t_{n+1} \to t_{n+}$, 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 extreme shift 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.
- Which 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.
- Which is the value of the forecasting error $e_{t_{n+1}}$.

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 the analyzed 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 Sections 3.5 and 3.6 for a real data example and a simulation study.

Let us notice that the other methods for 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 DES for irregular data, see Cipra (2006) or Section 2.1, and the method based on the DLS estimation of linear trend, see Hanzák and Cipra (2008) or Section 2.1. But these methods are less flexible since they use only one smoothing constant instead of two as 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.

3.4 Suggested solution

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

The first possibility is to modify somehow the analyzed time series before we apply the forecasting method to it. One could go through the series and find all pairs of time-close observations in it (certain threshold on the time step length must be set). For such a pair, a reasonable solution is to transform 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 disadvantage. The measurement noise variance of the constructed 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 formulas.

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 (3.10) where the forecasting error $e_{t_{n+1}}$ is multiplied by the potentially infinite expression $\gamma_{t_{n+1}}\alpha_{t_{n+1}}/(t_{n+1}-t_n)$, see (3.13).

But the problem can be seen also from the point of view of the weighting formula (3.7). 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 effects than by the true change in the time series level. So the weight given to $\tilde{T}_{t_k}$ in (3.7) should be decreasing with decreasing value of $t_k - t_{k-1}$.

We suggest this weight to be equal to

$$\frac{t_k - t_{k-1}}{t_n - t_{n-1}}(1-\gamma)^{t_n - t_k} \tag{3.14}$$

(the weight for $k = n$ is still equal to 1) so the formula (3.7) will be replaced by

$$T_{t_n} = \gamma_{t_n} \sum_{j=0}^{\infty} \frac{t_{n-j} - t_{n-j-1}}{t_n - t_{n-1}}(1-\gamma)^{t_n - t_{n-j}} \tilde{T}_{t_{n-j}}\,, \tag{3.15}$$

where now, in contrast to (3.8),

$$\gamma_{t_n} = \left[\sum_{j=0}^{\infty} \frac{t_{n-j} - t_{n-j-1}}{t_n - t_{n-1}}(1-\gamma)^{t_n - t_{n-j}}\right]^{-1}. \tag{3.16}$$

It is obvious that the recurrent formula (3.5) for the slope T remains unchanged. Only a different smoothing coefficient γ_{t_n} is used here, see (3.16). A recurrent formula for $\gamma_{t_{n+1}}$ analogous to that in (3.6) can be derived as well:

$$\gamma_{t_{n+1}} = \frac{\gamma_{t_n}}{\gamma_{t_n} + \frac{t_n - t_{n-1}}{t_{n+1} - t_n}(1-\gamma)^{t_{n+1} - t_n}}\,. \tag{3.17}$$

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 (2.15) to change it into (3.17). The initial value for γ_{t_n}, following the Wright's concept of a fixed point, is the same as in (2.6):

$$\gamma_{t_0} = 1 - (1-\gamma)^q\,. \tag{3.18}$$

None of other parts of a practical implementation of the method is affected. Looking at formula (3.17), 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 (3.10) again. Now for $t_{n+1} \to t_{n+}$, it is

$$\frac{\gamma_{t_{n+1}} \alpha_{t_{n+1}}}{t_{n+1} - t_n} \to \frac{\gamma_{t_n} \alpha_{t_n}}{(t_n - t_{n-1})(\alpha_{t_n} + 1)} < \infty \tag{3.19}$$

so there is no more problem of rapidly shifted $T_{t_{n+1}}$.

The weights suggested in (3.15) can be justified in the following way. Let us take $\gamma = 0$ and only a finite versions of our smoothing formulas ($j = 0, \ldots, J$) for a moment. Then formula (3.15) (using also appropriate finite version of (3.16)) will turn into

$$T_{t_n} = \frac{L_{t_n} - L_{t_{n-J-1}}}{t_n - t_{n-J-1}} \tag{3.20}$$

which seems to be more reasonable than the analogous finite version of (3.7):

$$T_{t_n} = \frac{1}{J+1} \sum_{j=0}^{J} \frac{L_{t_{n-j}} - L_{t_{n-j-1}}}{t_{n-j} - t_{n-j-1}}\,. \tag{3.21}$$

The slope is no more estimated as a simple arithmetic average of the individual one-step slopes. Now it is calculated as the "overall" slope which can be expressed as a weighted average of the one-step slopes, the weight being the corresponding time step. Adding an exponential weighting to this weighted average, we come to the suggested modification of slope estimation for Holt method.

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, see Cipra (2006) or Section 2.1, and Holt-Winters method for time series with missing observations, see Cipra et al. (1995) or Section 2.1.

With exponential trend we can use exactly the same modified formula (3.17) as well. But now we are loosing the justification provided by (3.20) which works exactly in the case of classical linear trend only (unless we change the slope estimation from arithmetic to geometric weighted average).

In the case of damped linear trend formula (3.17) is replaced by its analogue

$$\gamma_{t_{n+1}} = \frac{\gamma_{t_n}}{\gamma_{t_n} + \frac{g(t_n - t_{n-1})}{g(t_{n+1} - t_n)}(1-\gamma)^{t_{n+1}-t_n}}\,, \tag{3.22}$$

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

Seasonality in Holt-Winters method, either additive or multiplicative, has no impact on slope updating and so the suggested modification needs no special comments. It is used inside Holt-Winters in Chapter 4.

3.5 Real data example

The real data example used in this section is inspired by Wright (1986) who used a time series of men world record times for 1 mile run appr. from 1860 to 1970. The observation time was the year in which the record was established. When a new record time was established multiple times during one calendar year, just the final (best) time was considered.

We will use a similar time series with the difference that the observation times are recorded with a day by day precision. It covers the IAAF era, starting with time 4:14.4 by John Paul Jones (USA) run on 31st May 1913 in Allston. The latest record time is 3:43.13 by Hicham El Guerrouj (Morocco) run on 7th July 1999 in Rome[1]. Totaly the series contains 32 observations with average time spacing of 2.78 years. The observation times are expressed in years (as fractional numbers) and the record times in seconds (with precision of 0.1 and later 0.01 second).

Primarily we will apply the Holt method by Wright (1986) and its modified version suggested in Section 3.4 to smooth and forecast this time series. Fixed smoothing constants $\alpha = 0.2$ and $\gamma = 0.02$ have been used. Low value of γ has been used to reflect the relatively stable linear trend.

The initial values L_0 and T_0 have been constructed by fitting a linear regression through the first 10 observations of the series. DLS with weights decreasing towards future with discount factor $1 - \sqrt{\alpha\,\gamma}$ have been used, see Hanzák (2007).

In addition to this, the DES and the method of DLS linear trend described in Section 2.2 will be applied. Here the smoothing constant α will be optimized according to RMSE of one-step-ahead forecasting errors. All the computations were done in author's own application DMITS, see Chapter 7.

The results are as follows. DES and DLS liner trend achieved RMSE of 1.2626 and 1.2680 respectively, with optimal α of 0.0994 and 0.1111 respectively. The modified Holt method achieved RMSE of 1.2552, i.e. it slightly outperformed the first two methods. It has an advantage of two independent smoothing constants (even that they were not optimized but set fixed) so it can smooth the series slope more than its level.

The original version of Holt method provided by Wright (1986) absolutely failed - it achieved RSME of 2.8470, i.e. more than twice higher than all the previous three tested methods. And this happened despite the fact that we chose very small γ value which should mitigate the potential impact of time-close observations.

[1] http://en.wikipedia.org/wiki/Mile_run_world_record_progression. Accessed on 9th February 2014.

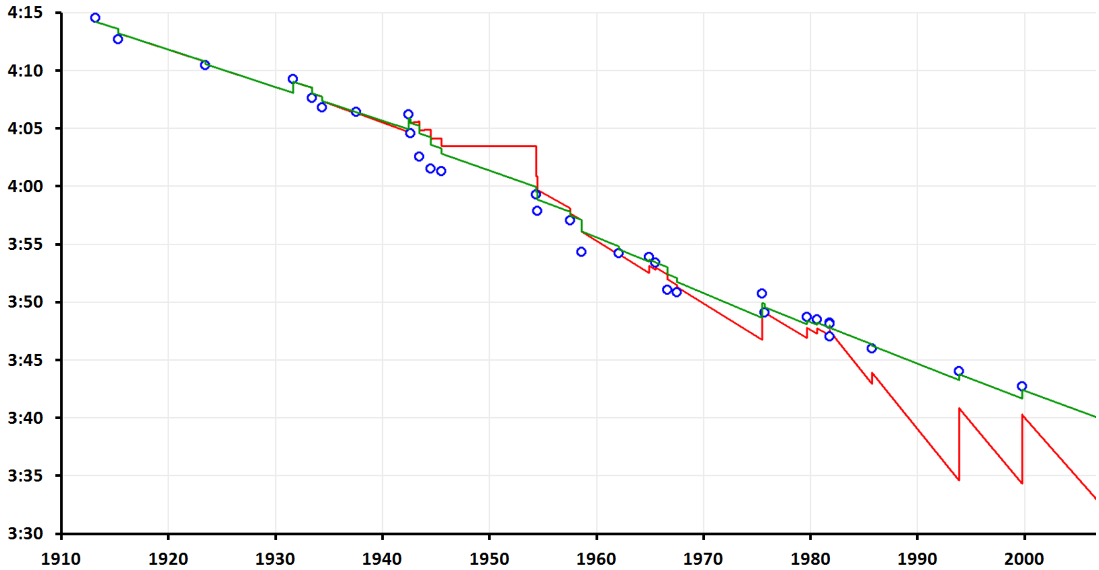

Figure 3.1: One mile run men world record times [M:SS], 1913-1999. Holt method with fixed $\alpha = 0.2$, $\gamma = 0.02$ (original method - red, modified method - green).

Let us have a look in detail on Figure 3.1 what happens in comparison of the two versions of Holt method. The version by Wright (1986) is plotted in red color, our modified version in green. We can see two serious problems of the original method. The first one due to the record time 4:06.2 run equally on 1 July 1942 and on 10 July 1942. This deviated the slope estimate upwards and caused significant forecasting errors in the following period (then it turned slightly in the opposite direction on 21 June 1954). The second problem occurred due to the record times 3:48.53, 3:48.40 and 3:47.33 run between 19 and 28 August 1981. This deviated the slope estimate downwards and caused significant forecasting errors for the remaining part of the series.

Slope smoothing coefficient γ_{t_n} ranged from 0.045 to 0.066 over time in the original version of the method, with variation coefficient just of 9 %. In the modified method, γ_{t_n} ranged from 0.00011 to 0.17 over time, with variation coefficient of 97 %. It shows how big the influence of the modification of formula for γ_{t_n} is when the time steps are very variable (they have variation coefficient of 101 % in this case).

3.6 Simulation study

In this section we will evaluate what is the improvement of a forecasting accuracy due to the suggested modification of Holt method using a simulation study. Both the original and modified Holt method were implemented in author's own application DMITS, see Chapter 7.

We will evaluate this improvement depending on the intensity of time-close observations presence in data. This *intensity* means

1. *frequency* of the time-close observations in data, i.e. how often the problem occurs in the series,

2. *closeness* of the time-close observations, i.e. how dramatic the problem is.

In addition we try three levels of *smoothness* of the series. Totaly we use 21 different simulated time series.

First 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, see Table 3.1. Finally we normalized the time axis of the resulting irregular time series so that its average time spacing q is equal to 1.

Closeness	**Frequency**	**Time step distribution**	***Q***
-	-	$\mathsf{U}\{1, 2, 3, 4\}$	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

Table 3.1: Used time step distributions. U is a discrete uniform distribution and δ_x is the Dirac measure at point x.

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 $L_0 = T_0 = 0$ and the one-step-ahead forecasting errors were independent identically distributed with $\mathsf{N}(0, 1)$. All the generated irregular time series have 2000 observations[1].

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. Q is the expected value from Table 3.1, depending on the selected time step distribution. Using α^* and γ^* instead of α and γ to generate the regular series should guarantee similar optimal values of smoothing constants used for the sampled irregular time series.

[1] Using very long time series is used here as an alternative to generating multiple series from the same scheme.

As far as the implementation of both Wright's original method and the modified one is concerned, the initial values of L 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 again) 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 $= \sqrt{\text{MSE}}$ are presented for each of 21 time series and both the original and modified method, see Table 3.2.

			Original method			**Modified method**		
Freq.	**Closeness**	**Smooth.**	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

Table 3.2: Optimal α and γ values and the achieved RMSE for all 21 simulated time series, for both the original and the modified method.

From Table 3.2 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 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 totaly out of the mainstream of observations. Then it takes a certain time to join it again. Sometimes 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.

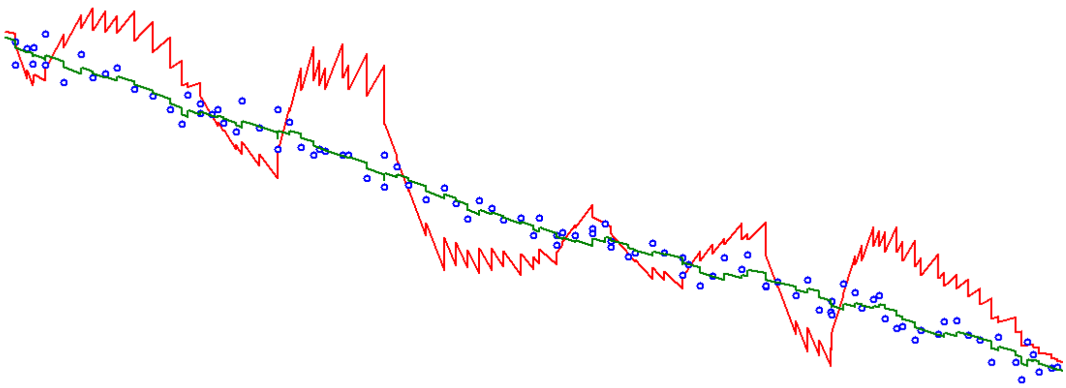

Figure 3.2: Illustration of Holt method with fixed $\alpha = 0.3$ and $\gamma = 0.1$ applied to a simulated time series (original method - red, modified method - green).

See Figure 3.2 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 3.2). 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, see e.g. Cipra (2008), p. 306. While the forecasts from the modified method are without problems (green line), those from the original one (red line) oscillates wildly.

3.7 Conclusion

Modification of Wright's version of Holt method for irregular time series suggested in this chapter 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.

Chapter 4

Holt-Winters method with general seasonality

4.1 Introduction

Holt-Winters method employs p seasonal indices (additive or multiplicative) to model the seasonal pattern of length p, see e.g. Holt (2004), Winters (1960) or Section 2.1. However, this brings couple of limitations. 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, $p = 12$). Especially this means that the number of observations per one period must be integer.

The calendar assignment is still possible in a time series with missing observations, see Cipra et al. (1995) or Section 2.1 for such an extension of Holt-Winters method. However, it is not possible in a general irregular time series and so there was no Holt-Winters method available for this case (see Bessa and Matos (2011)).

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. Holt-Winters method with general seasonality modeling is therefore suggested in this chapter. The goal is to offer a broader spectrum of possibilities for seasonality treatment while staying in the widely known and accepted framework of 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. Aldrin and Damsleth (1989), Ratinger (1996), Hyndman et al. (2002), Hanzák and Cipra (2008) or De Livera et al. (2011). In contrast to that, the method suggested in this chapter can be viewed as *ad hoc*, following the tradition of exponential weighting idea from Holt (2004), Winters (1960), Wright (1986), Cipra et al. (1995) or Hanzák (2008). This hopefully supports the understandability of the method while it does not harm its smoothing and forecasting performance; Aldrin and Damsleth (1989) and Hanzák and Cipra (2008) showed that the performance of *ad hoc* methods is fairly comparable with that of the optimal model-based ones.

In Section 4.2 Holt-Winters method with a general seasonality modeling (in its additive and multiplicative variants) is presented. The properties of this method are discussed, its theoretical justification based on *Discounted Least Squares* (DLS) estimation is given and the implementation details are outlined here. In Section 4.3 we discuss particular methods useful in practice, including linearly interpolated seasonal indices and trigonometric functions. Sections 4.4 and 4.5 compare the suggested methods numerically with the classical one on simulated and real data, respectively. Section 4.6 brings the conclusions of the chapter.

4.2 General seasonality modeling in Holt-Winters method

Seasonality can be generally modeled using $K \geq 1$ different real-valued functions $f_1, f_2, \dots, f_K$, 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:

$$S(t) = \sum\nolimits_{k=1}^{K} A^k f_k(t)\,, \tag{4.1}$$

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

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

$$\hat{y}_t = L_t + S(t) \tag{4.2}$$

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

$$\hat{y}_t = L_t \cdot \exp[S(t)]\,. \tag{4.3}$$

We suppose f_k just to be bounded. One can take functions f_k centered around 0 in a certain sense. It is also reasonable (but not necessary) for f_k to be linearly independent, see (4.1).

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 considered as well) and additive seasonality (the multiplicative case will be described later). We consider its level L_{t_n}, slope T_{t_n} and seasonal component

$$S_{t_n}(t) = \sum\nolimits_{k=1}^{K} A_{t_n}^k f_k(t) \tag{4.4}$$

at time t_n. Here $A_{t_n}^k$ 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 (2.22):

$$\hat{y}_{t_n+\tau}(t_n) = L_{t_n} + \tau T_{t_n} + S_{t_n}(t_n + \tau)\,, \tag{4.5}$$

$$\hat{y}_{t_n} = L_{t_n} + S_{t_n}(t_n)\,. \tag{4.6}$$

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 (2.32)-(2.34):

$$L_{t_{n+1}} = L_{t_n} + (t_{n+1} - t_n)\, T_{t_n} + \alpha_{t_{n+1}} e_{t_{n+1}}\,, \tag{4.7}$$

$$T_{t_{n+1}} = T_{t_n} + \gamma_{t_{n+1}} \alpha_{t_{n+1}} e_{t_{n+1}} / (t_{n+1} - t_n)\,, \tag{4.8}$$

$$A_{t_{n+1}}^k = A_{t_n}^k + \delta_{t_{n+1}}^k (1 - \alpha_{t_{n+1}}) e_{t_{n+1}} / f_k(t_{n+1})\,, \tag{4.9}$$

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 (4.9). Formulas (4.7) and (4.8) are equivalent to those in Wright (1986) and Cipra et al. (1995), see (2.32) and (2.33). The factor $\delta_{t_{n+1}}^k (1 - \alpha_{t_{n+1}})$ in (4.9) expresses the portion of $e_{t_{n+1}}$ which is absorbed to the kth seasonal component $A_{t_{n+1}}^k f_k(t_{n+1})$. The division by $f_k(t_{n+1})$ in (4.9) is due to (4.4) and it does not contradict to the *additive* seasonality used.

Smoothing coefficient $\alpha_{t_n} \in (0, 1)$ for level in (4.7) is updated in a recursive way, following the basic idea of exponential weighting, exactly as in Wright (1986) and Cipra et al. (1995):

$$\alpha_{t_{n+1}} = \frac{\alpha_{t_n}}{\alpha_{t_n} + (1 - \alpha)^{t_{n+1} - t_n}}\,, \tag{4.10}$$

where $\alpha \in (0,1)$ is a smoothing constant for level, see (2.5).

For the smoothing coefficient $\gamma_{t_n} \in (0,1)$ for slope in (4.8), we will use a modified updating formula (3.17) from Chapter 3:

$$\gamma_{t_{n+1}} = \frac{\gamma_{t_n}}{\gamma_{t_n} + \frac{t_n - t_{n-1}}{t_{n+1} - t_n}(1-\gamma)^{t_{n+1}-t_n}}\,, \tag{4.11}$$

where $\gamma \in (0,1)$ is a smoothing constant for slope. The modified coefficient $\gamma_{t_{n+1}}$ defined by (4.11) makes the slope estimate $T_{t_{n+1}}$ in (4.8) safe from a negative impact of the time distance $t_{n+1} - t_n$ being close to zero.

Smoothing coefficients $\delta^k_{t_n}$, $k = 1, \ldots, K$, for the seasonal amplitudes in (4.9) 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

$$W^k_{t_n} \equiv \sum\nolimits_{j=0}^{+\infty} (1-\delta^k)^{t_n - t_{n-j}} f_k^2(t_{n-j})\,. \tag{4.12}$$

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

$$W^k_{t_{n+1}} = (1-\delta^k)^{t_{n+1}-t_n} W^k_{t_n} + f_k^2(t_{n+1})\,. \tag{4.13}$$

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

$$\Delta^k_{t_{n+1}} \equiv f_k^2(t_{n+1}) \,/\, W^k_{t_{n+1}} \tag{4.14}$$

(we take again $0/0 = 0$). Since according to (4.12) 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 (4.14) is consistent with the fundamental idea of exponential weighting, see Wright (1986) for simple exponential smoothing. In (4.12) together with (4.14), 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 respects 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. Later we will give additional justification for the choice in (4.12) and (4.14).

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

$$\Delta_{t_{n+1}} \equiv 1 - \prod_{k=1}^{K} \left(1 - \Delta^k_{t_{n+1}}\right) \in [0,1] \tag{4.15}$$

be the total portion of the error absorbed instead of

$$D_{t_{n+1}} \equiv \sum_{k=1}^{K} \Delta^k_{t_{n+1}} \geq 0\,. \tag{4.16}$$

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

$$\delta^k_{t_{n+1}} \equiv \frac{\Delta_{t_{n+1}}}{D_{t_{n+1}}} \Delta^k_{t_{n+1}} \in [0,1]\,,\ \ k = 1, \ldots, K \tag{4.17}$$

(again take $0/0 = 0$). The motivating interpretation of (4.15) vs. (4.16) is that we rather imagine independence than disjointness of the k events of absorbtion with probabilities $\Delta^k_{t_{n+1}}$.

Let us summarize that the suggested Holt-Winters method with general seasonality consists of formulas (4.7)-(4.11) and (4.13)-(4.17). In total, one needs to keep $4+2K$ numerical variables in memory which are updated over time by the above listed recursive formulas. The computational complexity of the method is comparable with that from Cipra et al. (1995) and is reduced with lower number K of seasonal functions f_k or when some of them are repeatedly equal to 0 (see Section 4.3 for concrete examples).

Method's properties and theoretical justification. The smoothing coefficients $\delta^k_{t_{n+1}}$ as defined in (4.12)-(4.17) have reasonable properties:

- By $A^k_{t_{n+1}}$ update we move from $\hat{y}_{t_{n+1}}(t_n)$ closer to $y_{t_{n+1}}$. Summing the k movements, we come to $S_{t_{n+1}}(t_{n+1}) = S_{t_n}(t_{n+1}) + \Delta_{t_{n+1}}(1 - \alpha_{t_{n+1}})e_{t_{n+1}}$, see (4.4), (4.9) and (4.15)-(4.17). Thus the total portion of $e_{t_{n+1}}$ absorbed by seasonals is $\Delta_{t_{n+1}}(1 - \alpha_{t_{n+1}}) \in [0,1]$. Compare this with (2.34).

- 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 (4.12) and (4.14).

- If $f_k(t_{n+1}) \to 0$ then (ceteris paribus) $\delta^k_{t_{n+1}}/f_k(t_{n+1}) \to 0$. This means that we do not need to worry about values of f_k near to 0, see the division in (4.9).

To justify the concrete choice in (4.12) and (4.14) for $\delta^k \equiv \delta$, let us consider a *Discounted Least Squares* (DLS) estimation of K parameters A^k in the linear regression model

$$y_t \approx \sum\nolimits_{k=1}^{K} A^k f_k(t) \tag{4.18}$$

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

$$\Sigma_n(\boldsymbol{A}) \equiv \sum_{j=0}^{\infty} \left[y_{t_{n-j}} - \sum\nolimits_{k=1}^{K} A^k f_k(t_{n-j}) \right]^2 (1-\delta)^{t_n - t_{n-j}} \,, \tag{4.19}$$

where we denoted $\boldsymbol{A} = (A^1, \ldots, A^K)'$. Let us stress that we do not consider the level-trend component $L + t\,T$ in (4.18) 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

$$\boldsymbol{A}_{t_n} = (\boldsymbol{F}'_n \boldsymbol{D}_n \boldsymbol{F}_n)^{-1} \boldsymbol{F}'_n \boldsymbol{D}_n \boldsymbol{Y}_n \,, \tag{4.20}$$

where

$$\boldsymbol{F}_n = \{ f_k(t_{n-j}) \}_{j=0,1,2,\ldots}^{k=1,\ldots,K} \tag{4.21}$$

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

$$\hat{y}_{t_{n+1}}(t_n) = \sum\nolimits_{k=1}^{K} A^k_{t_n} f_k(t_{n+1}) \tag{4.22}$$

the regression prediction of $y_{t_{n+1}}$ using the estimate $\boldsymbol{A}_{t_n}$ and

$$e_{t_{n+1}} = y_{t_{n+1}} - \hat{y}_{t_{n+1}}(t_n) \tag{4.23}$$

the corresponding prediction error. Since it is

$$\Sigma_{n+1}(\boldsymbol{A}) = (1-\delta)^{t_{n+1} - t_n} \Sigma_n(\boldsymbol{A}) + \left[y_{t_{n+1}} - \hat{y}_{t_{n+1}}(t_{n+1}) \right]^2 \,, \tag{4.24}$$

$y_{t_{n+1}} = \hat{y}_{t_{n+1}}(t_n)$ implies $\boldsymbol{A}_{t_{n+1}} = \boldsymbol{A}_{t_n}$[2]. This fact together with (4.20) gives us

$$\boldsymbol{A}_{t_{n+1}} = \boldsymbol{A}_{t_n} + (\boldsymbol{F}'_{n+1} \boldsymbol{D}_{n+1} \boldsymbol{F}_{n+1})^{-1} \{ f_k(t_{n+1}) \}_{k=1,\ldots,K} e_{t_{n+1}} \,, \tag{4.25}$$

[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 $\hat{y}_{t_{n+1}}(t_{n+1}) = \hat{y}_{t_{n+1}}(t_n) = y_{t_{n+1}}$.

where $\{f_k(t_{n+1})\}_{k=1,\dots,K}$ 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

$$f_k \circ f_l \equiv \sum\nolimits_{j=0}^{+\infty} (1-\delta)^{t_{n+1}-t_{n+1-j}} f_k(t_{n+1-j}) f_l(t_{n+1-j}) = 0 \tag{4.26}$$

for all $k \neq l$, we get

$$A^k_{t_{n+1}} = A^k_{t_n} + \frac{f_k(t_{n+1})}{W^k_{t_{n+1}}} e_{t_{n+1}} = A^k_{t_n} + \Delta^k_{t_{n+1}} e_{t_{n+1}} / f_k(t_{n+1}) \,, \tag{4.27}$$

where $W^k_{t_{n+1}}$ and $\Delta^k_{t_{n+1}}$ are defined in the same way as before. This result supports the definition of $\Delta^k_{t_{n+1}}$ in (4.14).

Ignoring the possible non-zero off-diagonal elements of matrix $\boldsymbol{F}'_{n+1}\boldsymbol{D}_{n+1}\boldsymbol{F}_{n+1}$ is the reason why we need to do the normalization in (4.17). If we solved correctly $K \times K$ matrix inversion in (4.25), 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 (4.12)-(4.17) based on the diagonality assumption. If the functions f_k are approximately orthogonal (i.e. their scalar products $f_k \circ f_l$ for $k \neq l$ are *almost* zero when compared to $f_k \circ f_k = W^k_{t_{n+1}}$) 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 $\{1, t, f_1(t), \dots, f_K(t)\}$ using DLS estimation method with a certain discount factor. However, even if it is not necessary to invert $(K+2) \times (K+2)$ matrices, we lose the important flexibility of having three independent smoothing constants as in Holt-Winters method. See Chapter 5 for more details on this topic.

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 (4.5) and (4.6) with

$$\hat{y}_{t_n+\tau}(t_n) = (L_{t_n} + \tau T_{t_n}) \exp\left[S_{t_n}(t_n+\tau)\right] \,, \tag{4.28}$$

$$\hat{y}_{t_n} = L_{t_n} \exp\left[S_{t_n}(t_n)\right] \,. \tag{4.29}$$

The recursive formula (4.9) for the amplitudes update is simply changed to

$$A^k_{t_{n+1}} = A^k_{t_n} + \delta^k_{t_{n+1}}(1-\alpha_{t_{n+1}}) \left[\ln y_{t_{n+1}} - \ln \hat{y}_{t_{n+1}}(t_n)\right] \,. \tag{4.30}$$

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 (4.28) and (4.29) we do the reverse conversion from additive to multiplicative.

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 4.3. Generally with higher K we are able to model more precisely even complicated patterns but we must beware of over-fitting. See Sections 4.4 and 4.5 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 $(0, 1]^3$ as in the case of the classical Holt-Winters method.

- 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 Chatfield and Yar (1988) 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 (4.12)

$$W_0^k \approx \sum_{j=0}^{+\infty} (1 - \delta^k)^{jq}\, \overline{f_k^2} = \frac{\overline{f_k^2}}{1 - (1 - \delta^k)^q}\,, \tag{4.31}$$

  where $\overline{f_k^2}$ is the average squared value of f_k over the available observation times and q is the average time spacing of the series.

4.3 Useful special cases

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 (2.22)-(2.25), we simply take $K = p$ and

$$f_k(t) = \begin{cases} 1 & \text{if } \ (t \mod p) = k\,, \\ 0 & \text{otherwise}\,. \end{cases} \tag{4.32}$$

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:

$$\delta_t^k = \begin{cases} 1 - (1-\delta)^p & \text{if } \ (t \mod p) = k \,, \\ 0 & \text{otherwise} \,. \end{cases} \tag{4.33}$$

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, see also Cipra et al. (1995).

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 treatment of multiplicative seasonal indices seems to be more reasonable and theoretically consistent, however the practical impact on the results is not significant due to local linearity of exponential and logarithm functions.

Normalized seasonal indices. In Chatfield and Yar (1988) 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 (4.32) with

$$f_k(t) = \begin{cases} 1 & \text{if } \ (t \mod p) = k \,, \\ -1/(p-1) & \text{otherwise} \,. \end{cases} \tag{4.34}$$

The functions $f_k(t)$ are now centered around 0 and so the whole seasonal component $S(t)$ defined in (4.4) is as well. They remain linearly independent and approximately orthogonal for $p \gg 0$. Now always all the amplitudes A^k are updated in a single time step.

Missing observations. By taking K, f_k and δ^k the same as for the classical Holt-Winters method 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 Cipra et al. (1995)[1]. Again only the single amplitude A^k belonging to the

[1] We just use the modified slope update robust to time-close observations.

actual calendar unit of t is updated in a single time step. However, 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.

Restricted seasonal indices. Let us still consider a regular time series or at worst a time series with missing observation. For a particular real time series, there is sometimes a good reason to assume that the p seasonal amplitudes are not independent of each other but there are some logical linear restrictions between them. Taking this into account we can reduce the number of seasonal parameters used and eventually obtain a better out-of-sample forecasting accuracy thanks to more robust estimation.

The restrictions can have various forms, for example:

- two (neighboring or not) seasonal amplitudes are equal,
- certain seasonal amplitude is an average of the two surrounding ones,
- certain seasonal amplitude is exactly half of another etc.

All these restrictions can be simply imposed on the seasonal pattern of Holt-Winters method by reducing the number K of seasonal functions f_k used, analogously as in linear regression models.

The examples of real situations where such seasonal restrictions could make sense can be found easily. E.g. some monthly time series with annual seasonality can have the same seasonal amplitude in July and August due to school holidays located in these two months. Or a time series of daily TV channel's audience shares can exhibit the same values on two different weekdays due to the same programming scheme of the respective TV channel (consider Tuesdays and Thursdays of Czech TV Nova).

Sometimes we observe historically that at a certain seasonal unit the seasonal pattern is linear and so the respective amplitude can be set equal to the average of the two neighboring ones (i.e. linearly interpolated by them). This means in fact using the same idea as in *sparse seasonal indices* introduced later in this section.

Sometimes the seasonal deviation in two consecutive months is caused by the same factor which effect is spread over these two months. Historically we can estimate the ratio between the two months and assume it to be constant over time. Then just the overall strength of the effect is estimated as one parameter driving the seasonal pattern in the both months.

Interpolated seasonal indices. To cover the inter-calendar observations and/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 independently of the period length $p \in (0, +\infty)$ and with a general time axis origin $o \in \mathbb{R}$. Let us define

$$f_k(t) = \left\{1 - \min_{j \in \mathbb{Z}} \left| \frac{K(t-o)}{p} - (jK + k) \right| \right\}^+ . \qquad (4.35)$$

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 $f_K(o) = 1$. See Figure 4.1 for illustration.

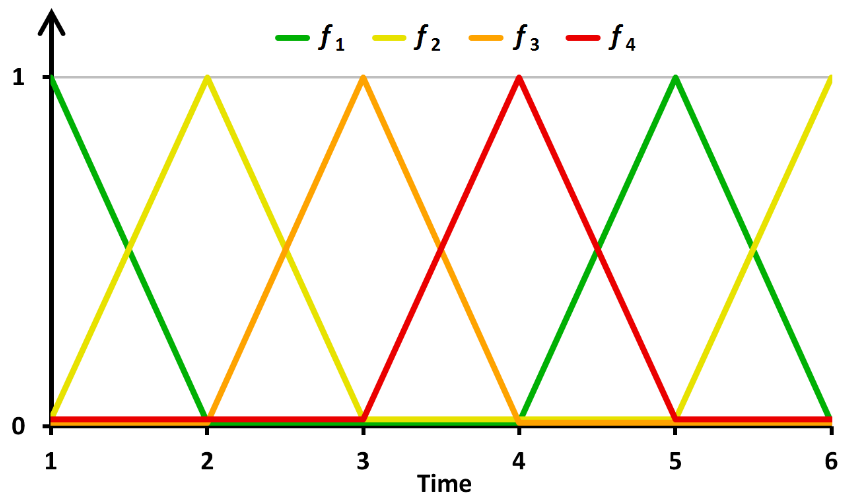

Figure 4.1: Interpolated seasonal indices with $K = 4$, $p = 4$ and $o = 0$.

By setting the amplitudes A^k we can form any p-periodic continuous $K-$piecewise linear seasonal pattern. The amplitudes A^k will appear as the pattern values at the K equidistant break points.

Functions f_k, $k = 1, \dots, K$ on $[0, p]$ interval are a special case of so called B-splines (**b**asis splines). In particular, they are the second order and the first degree B-splines, meaning that each function f_k is formed piecewise by two polynomials of the first degree (i.e. straight lines) and the connection is continuous, see Craig (2004). Higher order and degree B-splines are multiple piecewise polynomial curves of higher degrees with higher order of continuity (higher derivatives are continuous).

The 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 $\overline{f_k^2} = (2/3)/K$ 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 time series with period $p \in \mathbb{N}$ (see Sections 4.4 and 4.5):

- **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.

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

When one wants the formed seasonal pattern to be smooth (with continuous first derivative), the third order and the second degree B-splines can be used (instead of the second order and the first degree ones). These consist of 3 quadratic polynomials connected with continuous first derivative. The price one has to pay for this higher order continuity is that always two or three seasonal amplitudes are updated in each time step (not just one or two as in the lower order/degree case).

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 shift 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 now), we define

$$f_1(t) = \sin \frac{2\pi t}{p}\,, \quad f_2(t) = \cos \frac{2\pi t}{p}\,, \quad f_3(t) = \sin \frac{4\pi t}{p}\,, \quad f_4(t) = \cos \frac{4\pi t}{p}\,. \qquad (4.36)$$

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 $2h \leq p/q$ to prevent 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 $\overline{f_k^2} \equiv 1/2$.

Multiple seasonality. Half hourly electricity demand time series contains two different seasonalities: daily (period 48) and weekly (period $7 \cdot 48 = 336$). To construct forecasts, Taylor (2003) used a *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 be found e.g. in Taylor (2008).

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 (4.32), with $p_k = 48$ for the daily set and $p_k = 336$ for the weekly one.

Non-periodic functions f_k. Although the Holt-Winters method traditionally deals with seasonality, the general concept suggested in this chapter can be used even to cope with non-seasonal variations in time series values. It is enough that the variation can be attributed to a certain "regressor" function f quantifying certain factor influencing the time series values. Function f must be defined (observed) at the observation points of the analyzed time series y. It does not need to be periodic, yet it can be used within Holt-Winters method as suggested with no problems (the periodicity is not required by formulas (4.7)-(4.11) and (4.13)-(4.17)). To be eligible for predicting future values of the series y, values of f just must be known in advance for the future we are interested in. For example certain calendar effects can be captured by f or anything that is planned for sure into future (TV channel's programming scheme, dates of holding major sport events etc.)

4.4 Simulation study

In this section we will test the classical Holt-Winters method, the method with shifted seasonal indices and the method with trigonometric functions (see Section 4.3) on the simulated regular time series with locally constant trend and additive seasonality with period length $p = 7, 12$ and 24. All the three methods were implemented in author's own application DMITS, see Chapter 7.

The time series generating model used is

$$y_t = L_t + S_t + \varepsilon_t, \quad \varepsilon_t \sim iid\, \mathsf{N}(0,1), \tag{4.37}$$
$$L_t = L_{t-1} + \mu_t, \quad \mu_t \sim iid\, \mathsf{N}(0, 0.1^2) \tag{4.38}$$

with ε_t and μ_t mutually independent. The seasonal component S_t used in (4.37) follows

$$S_t = (1-\nu)(S_{t-1} + S_{t-p} - S_{t-p-1}) - \nu \sum_{j=t-p+1}^{t-1} S_j + \pi_t, \; \pi_t \sim iid\, \mathsf{N}(0,1), \tag{4.39}$$

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 (4.37)-(4.39) by $L_0 = 0$ and $S_j = 0$ for $j = -p, \ldots, 0$.

In SARIMA or state space models (see e.g. Ratinger (1996), Hyndman et al. (2002) or De Livera et al. (2011)) 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 corresponding seasonal pattern is not autocorrelated or smooth at all. But such a situation is rather rare in reality and thus the model (4.39) 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 (4.39). 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 (4.39) 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 (4.37) and (4.38), and a low fixed value of $\gamma = 0.05$ purposely to prevent from three-dimensional smoothing constants optimization.

		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

Table 4.1: Average out-of-sample RMSE and average ranking of the three methods tested on nine different simulation setups.

The average out-of-sample RMSE and the average ranking of the methods (1 = best, 3 = worst) are presented in Table 4.1. 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. Surprisingly the results 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 (4.39). 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. However, it is non-trivial to set up a neutral generating model useful for the comparison of the methods.

4.5 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 Hyndman (2010):

1. **AIR** - International 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 4.4, we will test also the "sparse indices" method, see Section 4.3. For **AIR**, **GAS** and **FLOW** we use a multiplicative seasonality, for **TEMP** and **LEVEL** the additive seasonality is used in all four methods. All the methods were implemented in author's application DMITS, see Chapter 7.

The smoothing constants α, γ and δ and (if needed) the number h of full harmonics are optimized with respect to RMSE based on the one-step-ahead forecasting errors through the whole series. The same in-sample RMSE values are reported in Table 4.2, 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

Table 4.2: Achieved minimal in-sample RMSE and autocorrelation ρ_e for five real time series and four methods tested.

"Shifted indices" method is the best one for all the 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 follows 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 (there is a major peak in summer but also another smaller one around Christmas and New Year and also in March, see Figure 4.2). The remaining three series have $h = 2$ or 3 as their optimum.

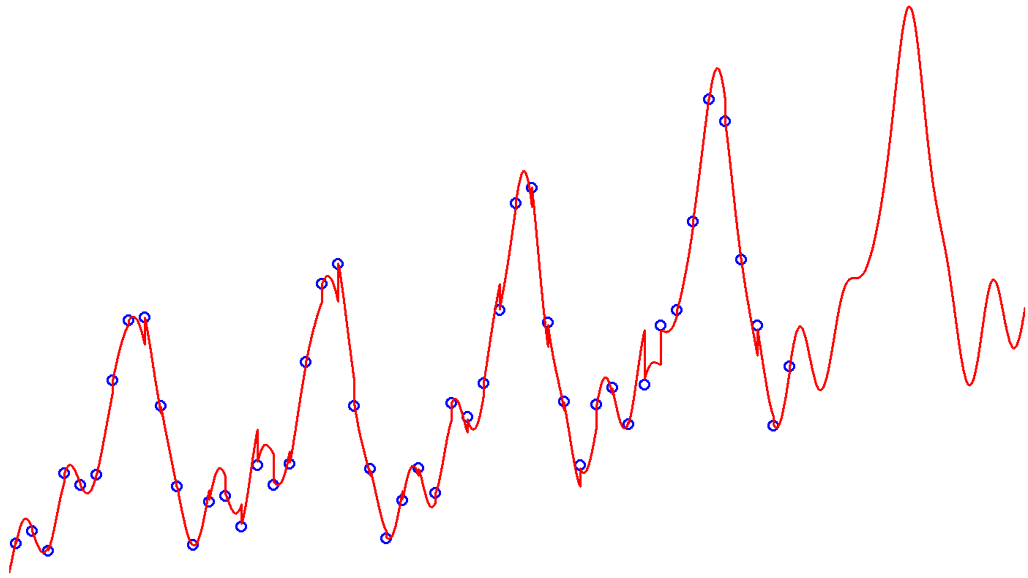

Figure 4.2: **AIR** series: multiplicative Holt-Winters method with five full harmonics.

See Figures 4.2, 4.3 and 4.4 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.

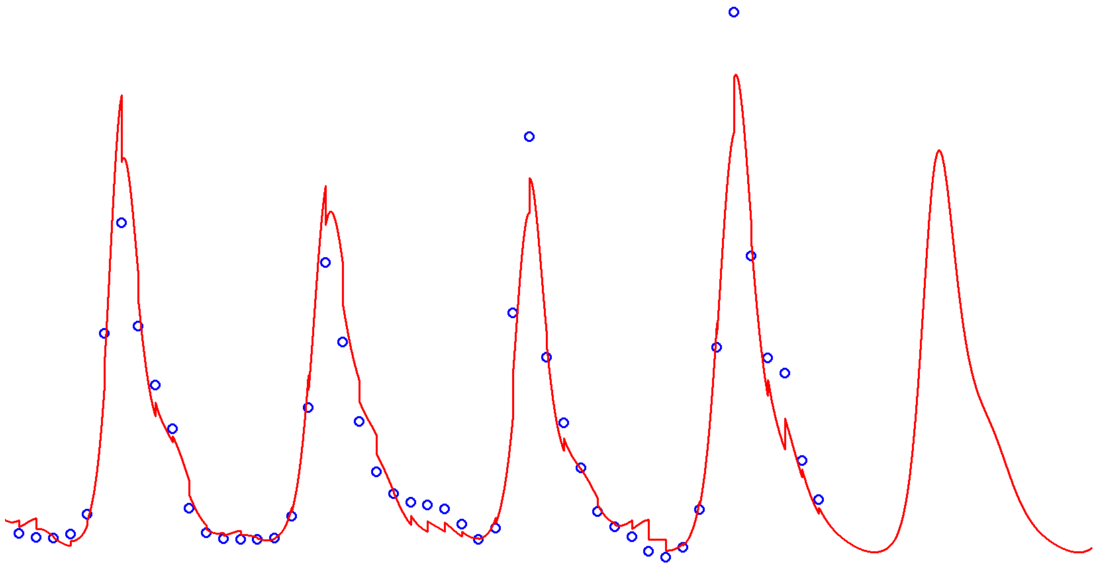

Figure 4.3: **FLOW** series: multiplicative Holt-Winters method with three full harmonics.

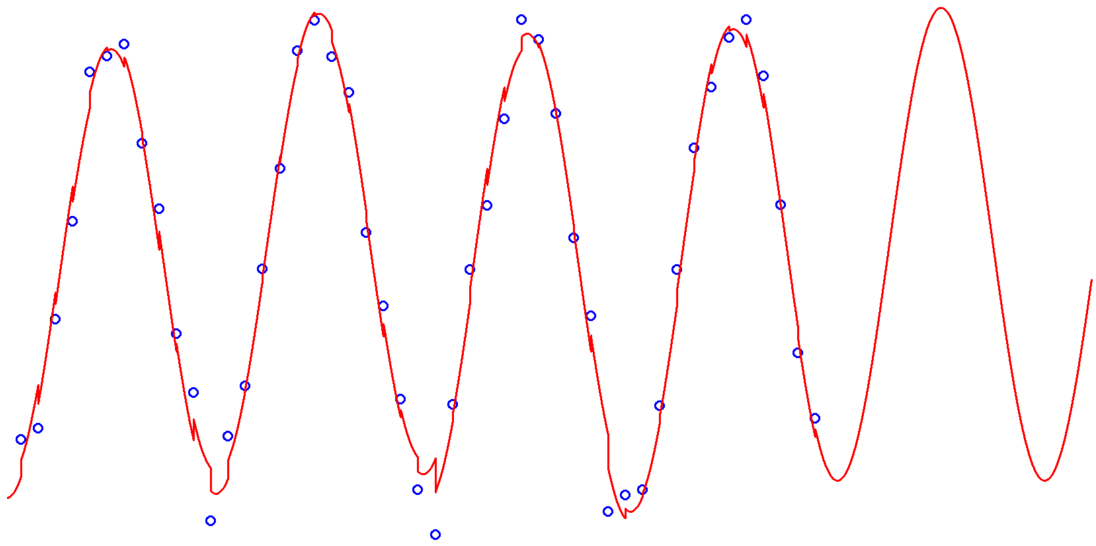

Figure 4.4: **TEMP** series: additive Holt-Winters method with one full harmonic.

4.6 Conclusion

General seasonality modeling concept was suggested in the framework of Holt-Winters method, both with additive and multiplicative seasonality. Several particular settings of seasonal functions used were suggested. General irregular time series can be handled, not just those with missing observations.

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 even for regular series it provides a relevant competitor to the classical Holt-Winters method.

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 be cautious 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. when 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.

For an example of a practical application of the suggested methods for irregular time series (forecasting electricity prices within an optimization models for electricity market agents), see Bessa and Matos (2012, 2014).

Chapter 5

DLS estimation and Holt-Winters method

5.1 Introduction

In Section 2.2 we presented Double Exponential Smoothing (DES, or Brown method) as a special case of Holt method. DES is derived via fitting a linear trend using a *Discounted Least Squares* (DLS) estimation with discount factor $\beta \in (0,1)$. Choice of a smoothing constant of DES, $\alpha = 1-\beta$, restricts us on a one-dimensional curve inside the unit square of Holt method smoothing constants α_H and γ_H, where it holds

$$\alpha_H = \alpha(2-\alpha) \quad \text{and} \quad \gamma_H = \frac{\alpha}{2-\alpha}\,. \tag{5.1}$$

Holt-Winters method is an extension of Holt method adding a seasonal component to it. Thus it works with three smoothing constants α_{HW}, γ_{HW} and δ_{HW}. Analogously as in the case of DES, we could try to estimate linear trend with seasonal dummies by DLS. This could possibly lead to a special case of Holt-Winters method with additive[1] seasonality, being again parameterized just by one smoothing constant $\alpha = 1-\beta$.

The purpose of this chapter is to investigate this possibility, i.e. derive the DLS method for seasonal time series, express the smoothing constants α_{HW}, γ_{HW} and δ_{HW} as functions of α and judge a practical usability of this method. Comparison with DES will be also done in terms of comparing relations between α_H and γ_H on one side and α_{HW} and γ_{HW} on the other side.

In the following text we will consider a regular time series y with locally linear trend and additive seasonality with period $p \geq 2$. The observations of y will be denoted as $\dots, y_{n-1}, y_n, y_{n+1}$, i.e. we assume an infinite history of y (similarly as

[1]Holt-Winters method with *multiplicative* seasonality of course can not be achieved by such a linear regression approach.

in the derivation of DES) up to time $n+1$. The smoothing constants notation will be used as in the paragraphs above.

5.2 DLS estimation of linear trend with seasonal dummies

Derivations in this section will start in an analogous way to Section 4.2 when we were finding a theoretical justification for the formulas of Holt-Winters method with general seasonality by assuming a DLS regression estimation.

Now we consider a linear regression model containing a linear trend and p seasonal dummies (indicators) $I_1, I_2, \ldots, I_p$:

$$y_t \approx T[t-(n+1)] + \sum\nolimits_{k=1}^{p} \sigma^k I_k(t)\,. \tag{5.2}$$

The unknown regression parameters are T (trend slope) and $\sigma^1, \sigma^2, \ldots, \sigma^p$ (seasonal levels). We replaced the intercept by putting all p (not just $p-1$) seasonal dummies into the regression formula. Let us suppose without loss of generality that the observation time $n+1$ belongs to the first seasonal dummy, i.e. that $I_1(n+1) = 1$.

The model parameters will be estimated using DLS with discount factor $\beta \in (0,1)$. The minimized DLS criterion based on the observations of y up to time n is

$$\Sigma_n(T,\boldsymbol{\sigma}) \equiv \sum_{j=0}^{\infty}\left[y_{n-j} + T(j+1) - \sum_{k=1}^{p}\sigma^k I_k(n-j)\right]^2 (1-\beta)^j\,, \tag{5.3}$$

where we denoted $(T,\boldsymbol{\sigma}) \equiv (T, \sigma^1, \sigma^2, \ldots, \sigma^p)'$. Denote the argument of minima of Σ_n by $(T_n, \boldsymbol{\sigma}_n) \equiv (T_n, \sigma_n^1, \sigma_n^2, \ldots, \sigma_n^p)'$. It is

$$(T_n, \boldsymbol{\sigma}_n) = (\boldsymbol{F}_n' \boldsymbol{D}_n \boldsymbol{F}_n)^{-1} \boldsymbol{F}_n' \boldsymbol{D}_n \boldsymbol{Y}_n\,, \tag{5.4}$$

where

$$\boldsymbol{F}_n = \begin{pmatrix} -1 & I_1(n) & I_2(n) & \ldots & I_p(n) \\ -2 & I_1(n-1) & I_2(n-1) & \ldots & I_p(n-1) \\ -3 & I_1(n-2) & I_2(n-2) & \ldots & I_p(n-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \end{pmatrix} \tag{5.5}$$

is the regression design matrix (with $p+1$ columns and infinite number of rows),

$\boldsymbol{D}_n = \mathrm{Diag}\{1, \beta, \beta^2, \ldots\}$ is an infinite dimensional diagonal discounting matrix and $\boldsymbol{Y}_n = (y_n, y_{n-1}, y_{n-2}, \ldots)'$ is an infinite dimensional column vector. Notice that the elements of $\boldsymbol{F}_n$, $\boldsymbol{D}_n$ and $\boldsymbol{Y}_n$ are stacked in the opposite order than usually, i.e. from the latest observation time n back into the history. The infinite dimensions of the matrices $\boldsymbol{F}_n$, $\boldsymbol{D}_n$ and $\boldsymbol{Y}_n$ turns the scalar products from finite sums to series sums which are (or can be assumed to be) convergent due to exponential decay of $\boldsymbol{D}_n$ elements.

Elements of a $(p+1)$-dimensional column vector $\boldsymbol{F}'_n \boldsymbol{D}_n \boldsymbol{Y}_n$ could be expressed using values of smoothing statistics of y in a similar way as in the case of DES. The first and second smoothing statistics $S^{[1]}$ and $S^{[2]}$ are the same as in (2.41). In addition we need to record also the seasonal pattern of the series. This can be done by using "seasonal" smoothing statistics $S^1, S^2, \ldots, S^p$ defined in a similar way as $S^{[1]}$ but always using just observations of y from the corresponding seasonal unit.

Now let us denote

$$\hat{y}_{n+1}(n) = \sum_{k=1}^{p} \sigma_n^k I_k(n+1) \tag{5.6}$$

the regression prediction of y_{n+1} using the estimate $(T_n, \boldsymbol{\sigma}_n)$ and

$$e_{n+1} = y_{n+1} - \hat{y}_{n+1}(n) \tag{5.7}$$

the corresponding prediction error. DLS criterion $\Sigma_{n+1}(T, \boldsymbol{\sigma})$ can be decomposed as

$$\Sigma_{n+1}(T, \boldsymbol{\sigma}) = \beta \Sigma_n(T, \boldsymbol{\sigma}) + [y_{n+1} - \hat{y}_{n+1}(n+1)]^2 \ . \tag{5.8}$$

In case that $y_{n+1} = \hat{y}_{n+1}(n)$ (i.e. the prediction error e_{n+1} is 0), it is $\hat{y}_{n+1}(n+1) = \hat{y}_{n+1}(n)$ and $(T_{n+1}, \boldsymbol{\sigma}_{n+1}) = (T_n, \boldsymbol{\sigma}_n)$, i.e. the regression parameters do not change after a new observation y_{n+1} is incorporated into the estimation ($(T_n, \boldsymbol{\sigma}_n)$ minimizes both summands on the right hand side of (5.8)).

This fact together with (5.4) gives us

$$(T_{n+1}, \boldsymbol{\sigma}_{n+1}) = (T_n, \boldsymbol{\sigma}_n) + (\boldsymbol{F}'_{n+1} \boldsymbol{D}_{n+1} \boldsymbol{F}_{n+1})^{-1} (0, e_{n+1}, 0, \ldots, 0)' \tag{5.9}$$

(we used the assumed fact that $I_1(n+1) = 1$). Thus all the estimates T and $\sigma^1, \sigma^2, \ldots, \sigma^p$ are updated by certain multiples of the prediction error e_{n+1} and the corresponding coefficients are rowed in the second column (or row) of the symmetric matrix $(\boldsymbol{F}'_{n+1} \boldsymbol{D}_{n+1} \boldsymbol{F}_{n+1})^{-1}$.

All the elements of the symmetric matrix $\boldsymbol{F}'_{n+1} \boldsymbol{D}_{n+1} \boldsymbol{F}_{n+1}$ are sums of infinite geometric, arithmetico-geometric or "quadratico-geometric" series so it is possible to express them explicitly. We make use of the following well known formulas

$$\sum_{j=0}^{\infty} \beta^j = \frac{1}{1-\beta}\,, \quad \sum_{j=0}^{\infty} j\beta^j = \frac{\beta}{(1-\beta)^2} \quad \text{and} \quad \sum_{j=0}^{\infty} j^2\beta^j = \frac{\beta(1+\beta)}{(1-\beta)^3} \tag{5.10}$$

and express the $(p+1) \times (p+1)$ matrix $\boldsymbol{F}'_{n+1}\boldsymbol{D}_{n+1}\boldsymbol{F}_{n+1}$ as

$$\begin{pmatrix} \frac{\beta(1+\beta)}{(1-\beta)^3} & \frac{-p\beta^p}{(1-\beta^p)^2} & -\beta\frac{(p-1)\beta^p+1}{(1-\beta^p)^2} & \dots & -\beta^{p-2}\frac{2\beta^p+p-2}{(1-\beta^p)^2} & -\beta^{p-1}\frac{\beta^p+p-1}{(1-\beta^p)^2} \\ \frac{-p\beta^p}{(1-\beta^p)^2} & \frac{1}{1-\beta^p} & 0 & \dots & 0 & 0 \\ -\beta\frac{(p-1)\beta^p+1}{(1-\beta^p)^2} & 0 & \frac{\beta}{1-\beta^p} & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ -\beta^{p-2}\frac{2\beta^p+p-2}{(1-\beta^p)^2} & 0 & 0 & \ddots & \frac{\beta^{p-2}}{1-\beta^p} & 0 \\ -\beta^{p-1}\frac{\beta^p+p-1}{(1-\beta^p)^2} & 0 & 0 & \dots & 0 & \frac{\beta^{p-1}}{1-\beta^p} \end{pmatrix}. \tag{5.11}$$

The matrix contains zeros in all cells except the first row, first column and main diagonal; it is co called (symmetric) *arrowhead* matrix. To calculate its inverse, we make use of the following lemma (formulation and proof by the author):

Lemma (symmetric arrowhead matrix inversion): Let

$$\boldsymbol{A} = \begin{pmatrix} a_1 & a_2 & a_3 & \dots & a_{n-1} & a_n \\ a_2 & b_2 & 0 & \dots & 0 & 0 \\ a_3 & 0 & b_3 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ a_{n-1} & 0 & 0 & \ddots & b_{n-1} & 0 \\ a_n & 0 & 0 & \dots & 0 & b_n \end{pmatrix} \tag{5.12}$$

be a real-valued symmetric arrowhead $n \times n$ matrix with $b_k \neq 0$ for all $k = 2, 3, \dots, n$.

Then it holds

$$\det\boldsymbol{A} = m_1 \prod_{k=2}^{n} b_k\,, \tag{5.13}$$

where

$$m_1 \equiv a_1 - \sum_{k=2}^{n} \frac{a_k^2}{b_k}\,. \tag{5.14}$$

If $\boldsymbol{A}$ is regular, i.e. $m_1 \neq 0$, then it is

$$\boldsymbol{A}^{-1} = \boldsymbol{L}'\boldsymbol{M}^{-1}\boldsymbol{L}\,, \tag{5.15}$$

where matrices $\boldsymbol{L}$ and $\boldsymbol{M}$ are defined as

$$\boldsymbol{L} = \begin{pmatrix} 1 & -\frac{a_2}{b_2} & -\frac{a_3}{b_3} & \dots & -\frac{a_{n-1}}{b_{n-1}} & -\frac{a_n}{b_n} \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix} \tag{5.16}$$

and $\boldsymbol{M} = \text{Diag}\left(m_1, \frac{1}{b_2}, \frac{1}{b_3}, \dots, \frac{1}{b_{n-1}}, \frac{1}{b_n}\right)$. The second row of matrix $\boldsymbol{A}^{-1}$ is

$$\left(-\frac{a_2}{b_2 m_1}, \frac{1}{b_2} + \frac{a_2^2}{b_2^2 m_1}, \frac{a_2 a_3}{b_2 b_3 m_1}, \dots, \frac{a_2 a_{n-1}}{b_2 b_{n-1} m_1}, \frac{a_2 a_n}{b_2 b_n m_1}\right). \tag{5.17}$$

Proof: Determinant of $\boldsymbol{A}$ is expressed directly following the definition of matrix determinant, benefiting from the fact that $\boldsymbol{A}$ is a sparse matrix.

Matrix $\boldsymbol{L}$ expresses the row operations done with $\boldsymbol{A}$ (and the same operations done also with its columns) that transform matrix $\boldsymbol{A}$ to a diagonal matrix $\boldsymbol{M}$, i.e. it is $\boldsymbol{L}\boldsymbol{A}\boldsymbol{L}' = \boldsymbol{M}$ (this can be verified directly). So it is $\boldsymbol{A} = \boldsymbol{L}^{-1}\boldsymbol{M}\boldsymbol{L}'^{-1}$ and thus $\boldsymbol{A}^{-1} = \boldsymbol{L}'\boldsymbol{M}^{-1}\boldsymbol{L}$. The second row of matrix $\boldsymbol{A}^{-1}$ can be directly calculated now. □

Let us denote the second column of matrix $(\boldsymbol{F}'_{n+1}\boldsymbol{D}_{n+1}\boldsymbol{F}_{n+1})^{-1}$ by $\left(\Delta^T, \Delta^1, \Delta^2, \dots, \Delta^p\right)'$. Using this notation, according to (5.9) we have

$$T_{n+1} = T_n + \Delta^T e_{n+1}\,, \tag{5.18}$$

$$\sigma^k_{n+1} = \sigma^k_n + \Delta^k e_{n+1}\,, \quad k = 1, 2, \dots, p\,. \tag{5.19}$$

Using the above stated Lemma, we can express $\left(\Delta^T, \Delta^1, \Delta^2, \dots, \Delta^p\right)'$ in terms of p and β. The most complicated part of the calculation is to express the auxiliary quantity m_1. During its calculation one needs to use formulas for *finite* sums of geometric, arithmetico-geometric or "quadratico-geometric" series:

$$\sum_{k=0}^{n} \beta^k = \frac{1-\beta^{n+1}}{1-\beta}\,, \tag{5.20}$$

$$\sum_{k=1}^{n} k\beta^k = \beta\frac{1-(n+1)\beta^n + n\beta^{n+1}}{(1-\beta)^2}\,, \tag{5.21}$$

$$\sum_{k=1}^{n} k^2\beta^k = \beta\frac{1+\beta-(n+1)^2\beta^n + (2n^2+2n-1)\beta^{n+1} - n^2\beta^{n+2}}{(1-\beta)^3}\,. \tag{5.22}$$

After certain algebraic manipulations we get

$$m_1 = \frac{p^2 \beta^p}{(1-\beta)(1-\beta^p)^2} \,. \tag{5.23}$$

Obviously $m_1 > 0$ and so according to the Lemma, the matrix $(\boldsymbol{F}'_{n+1} \boldsymbol{D}_{n+1} \boldsymbol{F}_{n+1})$ is regular (what does not surprise us). Formulas for $\left(\Delta^T, \Delta^1, \Delta^2, \ldots, \Delta^p\right)'$ now follow relatively easily:

$$\Delta^T = \frac{(1-\beta)(1-\beta^p)}{p} \,, \tag{5.24}$$

$$\Delta^1 = 1 - \beta^{p+1} \,, \tag{5.25}$$

$$\Delta^k = \frac{\left[(p+1-k)\beta^k + k - 1\right](1-\beta)}{p} \quad \text{for } k = 2, \ldots, p \,. \tag{5.26}$$

We can see that all these quantities lie in $(0, 1)$ interval which confirms their meaning as proportions of prediction error e_{n+1} being absorbed by individual regression parameters.

5.3 Smoothing constants of Holt-Winters via DLS

In this section we proceed to express the relation of the DLS regression method elaborated in previous Section 5.2 to the Holt-Winters method with additive seasonality. This means:

1. Answer the question whether the DLS approach leads to a special case of Holt-Winters method or not.

2. If so (at least approximately), express the values of corresponding smoothing constants α_{HW}, γ_{HW} and δ_{HW} in terms of p and β.

One can see from (5.24)-(5.26) that not just the slope estimate T, overall series level and seasonal index belonging to observation y_{n+1} (i.e. σ^1) are updated when moving from time n to time $n+1$. Also all the remaining seasonal indices $\sigma^2, \ldots, \sigma^p$ change their values relatively to each other (it is not $\Delta^2 = \ldots = \Delta^p$). So strictly speaking, DLS estimation does not lead to a special case of additive Holt-Winters method. However, the differences among $\Delta^2, \ldots, \Delta^p$ are relatively small when compared to Δ^1. Therefore *approximately* this method can be viewed as a special case of additive Holt-Winters method.

Holt-Winters method with additive seasonality can be formulated in its error-correction form as

$$L_{n+1} = L_n + T_n + \alpha_{HW} e_{n+1}\,, \tag{5.27}$$

$$T_{n+1} = T_n + \gamma_{HW} \alpha_{HW} e_{n+1}\,, \tag{5.28}$$

$$S_{n+1} = S_{n+1-p} + \delta_{HW}(1-\alpha_{HW}) e_{n+1}\,, \tag{5.29}$$

where L_n is level, T_n slope and S_n the seasonal index of the series y at time n, $\alpha_{HW}, \gamma_{HW}, \delta_{HW} \in (0,1)$ are the respective smoothing constants and e_{n+1} is still the one-step-ahead forecasting error.

We want first to express the smoothing constants α_{HW}, γ_{HW} and δ_{HW} in terms of $\Delta^T, \Delta^1, \Delta^2, \ldots, \Delta^p$, based on a comparison of (5.18) and (5.19) with (5.27)-(5.29).

Since L is a series level, its change corresponds to an average change of $p-1$ seasonal dummies σ^k, $k = 2, \ldots, p$. This means that we take

$$\alpha_{HW} = \frac{1}{p-1} \sum\nolimits_{k=2}^{p} \Delta^k\,. \tag{5.30}$$

Situation regarding T is clear - it must be

$$\gamma_{HW} = \Delta^T / \alpha_{HW}\,. \tag{5.31}$$

Change of S in Holt-Winters method can be viewed as a change of the corresponding seasonal index when compared to the average of the remaining $p-1$ seasonal indices (being not updated at the moment). So we should take

$$\delta_{HW} = \frac{\Delta^1 - \alpha_{HW}}{1-\alpha_{HW}}\,. \tag{5.32}$$

Now let us substitute (5.24)-(5.26) into (5.30)-(5.32) receiving

$$\alpha_{HW} = \frac{(1+\beta^p)(1-\beta)}{2} = \frac{\alpha}{2}[1+(1-\alpha)^p]\,, \tag{5.33}$$

$$\gamma_{HW} = \frac{2}{p} \cdot \frac{1-\beta^p}{1+\beta^p} = \frac{2}{p} \cdot \frac{1-(1-\alpha)^p}{1+(1-\alpha)^p}\,, \tag{5.34}$$

$$\delta_{HW} = \frac{1-\beta^p}{1-\frac{\beta^p(1-\beta)}{1+\beta}} = \frac{1-(1-\alpha)^p}{1-\frac{\alpha(1-\alpha)^p}{2-\alpha}}\,, \tag{5.35}$$

where $\alpha = 1-\beta$. Even though expressing α_{HW}, γ_{HW} and δ_{HW} in terms of p and α (instead of β) leads to slightly longer formulas, it is more convenient to do so because α_{HW}, γ_{HW} and δ_{HW} are increasing functions of α which is more natural to plot and comment. Let us now look in more detail at formulas (5.33)-(5.35).

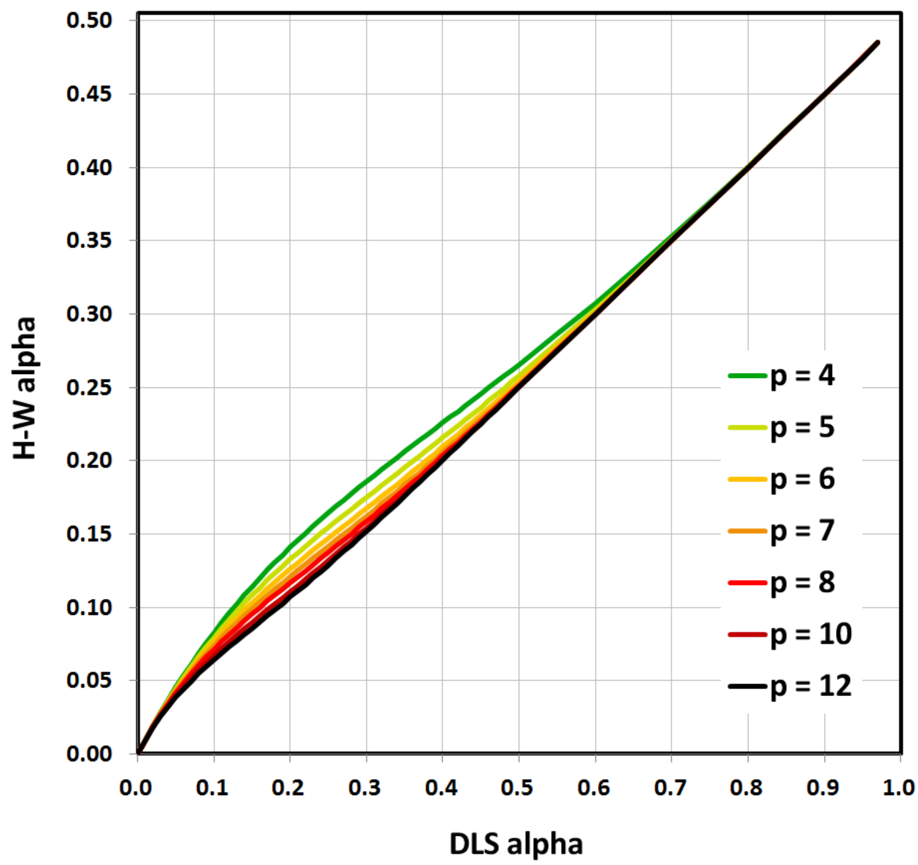

Figure 5.1: Additive Holt-Winters method by DLS regression: the dependence of smoothing constant α_{HW} on $\alpha = 1 - \beta$ and period length p.

Figure 5.1 shows how α_{HW} depends on α and p according to (5.33). Naturally α_{HW} is an increasing function of α. However, $\alpha_{HW} \to 0.5$ as $\alpha \to 1$, so α_{HW} is limited from above by 0.5, not by 1 as in Holt method. This is caused by the seasonal component which also absorbs part of the prediction error. The larger value of p, the higher portion of the error is absorbed by seasonal component (since there is longer time gap between neighboring observations from the same calendar unit) and the smaller α_{HW} is. But the dependency of α_{HW} on p is not extremely strong. For very small α, it is $\alpha_{HW} \approx \alpha$ (the slope of the curve at $\alpha = 0$ is 1), for larger α, it is $\alpha_{HW} \approx \alpha/2$ (the slope of the curve at $\alpha = 1$ is 0.5). In all cases it is $\alpha_{HW} \in [\alpha/2, \alpha]$. As $p \to \infty$, $\alpha_{HW} \to \alpha/2$ uniformly.

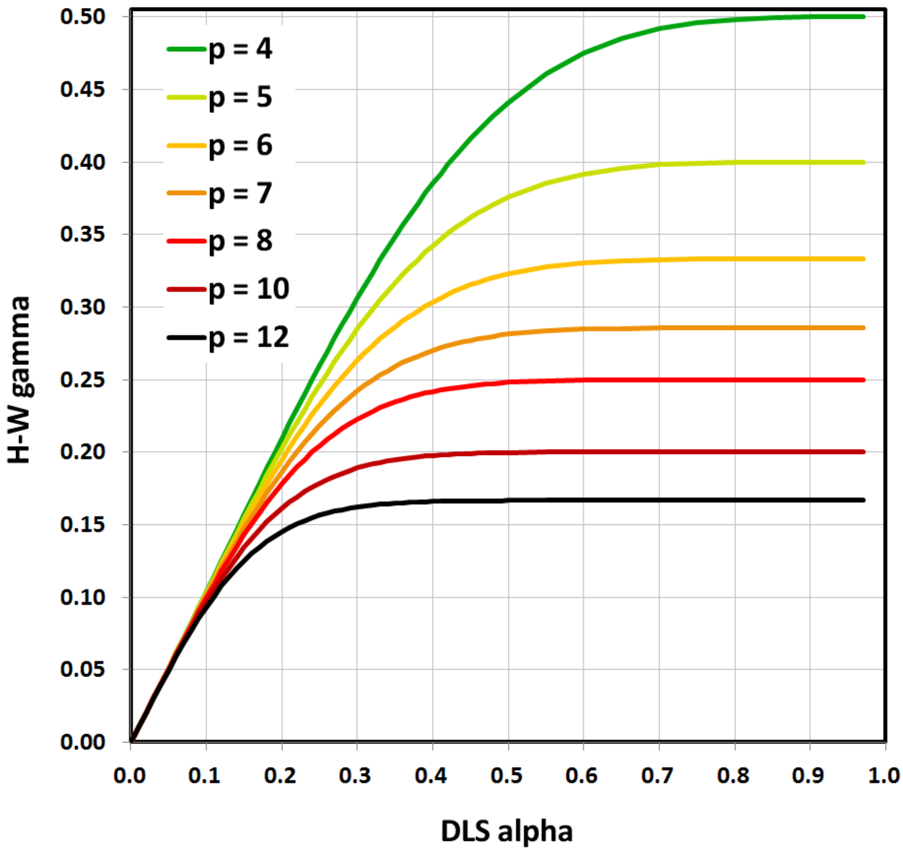

Figure 5.2: Additive Holt-Winters method by DLS regression: the dependence of smoothing constant γ_{HW} on $\alpha = 1 - \beta$ and period length p.

Figure 5.2 shows how γ_{HW} depends on α and p according to (5.34). Again, γ_{HW} is naturally an increasing function of α. The dependency is concave with horizontal asymptotic upper bound in the top right part of the graph. For small α, it is $\delta_{HW} \approx \alpha$ (the slope of the curve at $\alpha = 0$ is 1) and so especially it does not depend on p much. As $\alpha \to 1$, it is $\gamma_{HW} \to 2/p$. So again, γ_{HW} is limited from above not by 1 as in Holt method but by an upper bound lower then 1 which now depends strongly on p.

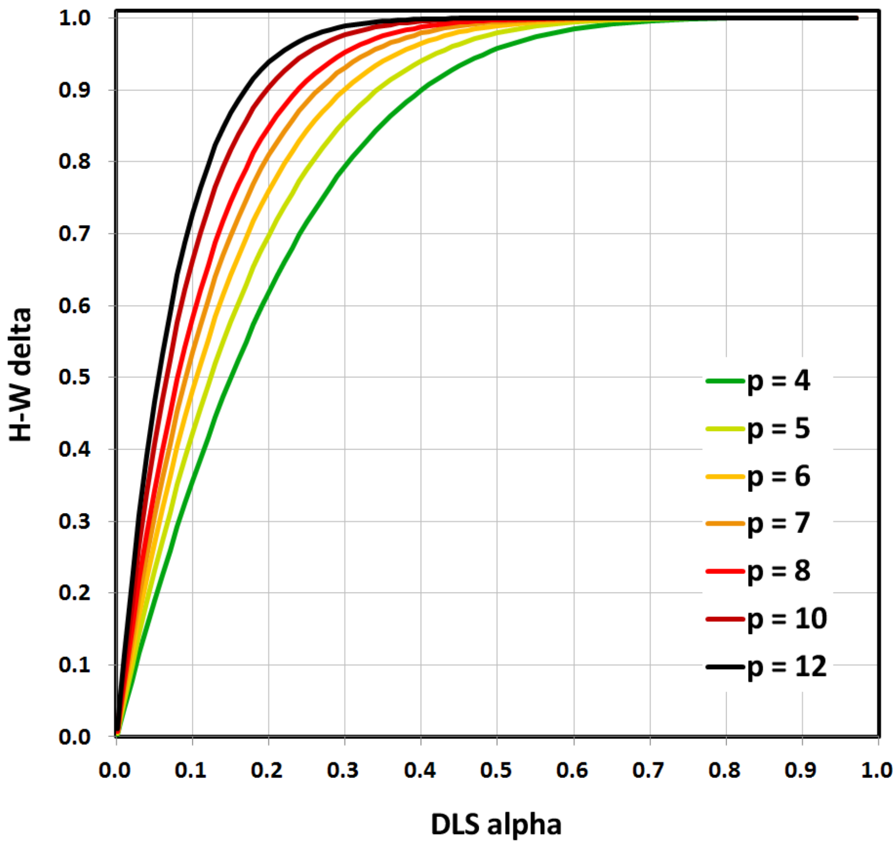

Figure 5.3: Additive Holt-Winters method by DLS regression: the dependence of smoothing constant δ_{HW} on $\alpha = 1 - \beta$ and period length p.

Figure 5.3 shows how δ_{HW} depends on α and p according to (5.35). Still δ_{HW} is an increasing and concave function of α. As $\alpha \to 1$, it is $\delta_{HW} \to 1$. This means that $\alpha_{HW} + (1 - \alpha_{HW})\delta_{HW} \to 1$ and so the whole prediction error is absorbed into the smoothed value of y as $\alpha \to 1$ (which clearly must be the case). There is a strong dependency of δ_{HW} on p - as was already explained, seasonal absorbtion is naturally higher for higher p. It holds

$$\delta_{HW} = \frac{1 - (1 - \alpha)^p}{1 - \frac{\alpha(1-\alpha)^p}{2-\alpha}} \approx 1 - (1 - \alpha)^p \tag{5.36}$$

approximately for all α and p and it is $\frac{\delta_{HW}}{1-(1-\alpha)^p} \to 1$ as $\alpha \to 0$ or $\alpha \to 1$. This expresses the natural exponential dependency of the effective seasonal smoothing constant on the period length p and the smoothing constant α per one time unit. See also (2.15) and (4.33) for similar δ conversions.

Notice the very high values of δ_{HW}. These are much higher than the ones usually used in practise: the seasonal pattern is typically quite stable over time (compared to the level or slope) so lower values of δ_{HW} use to be optimal. But the DLS regression estimation does not reflect this and makes quite significant updates of the seasonal parameters. This is the reason why DLS method for seasonal time series would provide typically significantly worse forecasting results than the general Holt-Winters method.

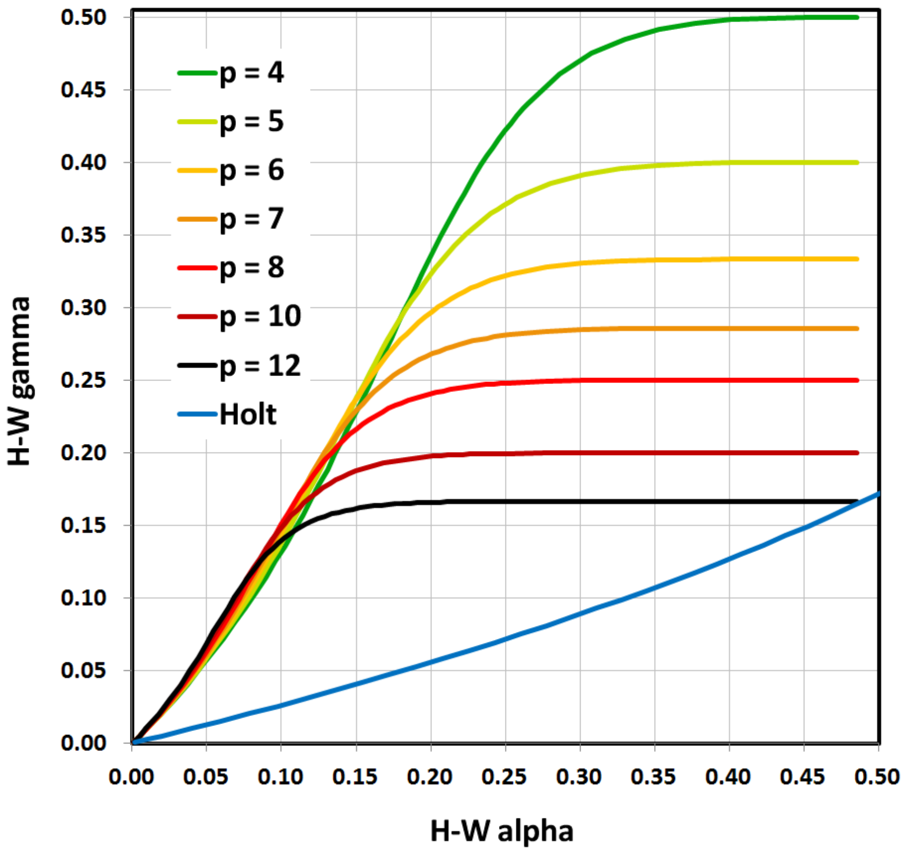

Figure 5.4: Additive Holt-Winters method by DLS regression: the dependence of smoothing constant γ_{HW} on smoothing constant α_{HW} and period length p.

Finally Figure 5.4 shows how γ_{HW} depends on α_{HW} and p. For comparison, the curve of dependency of γ_H on α_H in Holt method being restricted to DES (see (5.1)) was added to the same figure. As in Holt method, γ_{HW} is an increasing function of α_{HW}. The dependency is convex-concave with horizontal "asymptote" in the top right part of the graph. For very small α_{HW}, it is $\gamma_{HW} \approx \alpha_{HW}$ (and so especially it does not depend on p much). For moderate α_{HW} values (around 0.1), it is $\delta_{HW} \approx 1.4 \cdot \alpha_{HW}$. As $\alpha_{HW} \to 0.5$, it is $\gamma_{HW} \to 2/p$.

For small α_{HW} (up to 0.15), Holt-Winters method via DLS uses for a given value of α_{HW} appr. 4 to 5.5 times higher value of γ_{HW} than γ_H in Holt method with $\alpha_H = \alpha_{HW}$ (the ratio tends to 4 as $\alpha_{HW} = \alpha_H \to 0$, see (5.1)).

5.4 Conclusion

We showed that the DLS estimation of linear regression with linear trend and seasonal dummies is approximately equivalent to a certain special case of Holt-Winters method with additive seasonality. The dependency of the respective smoothing constants on $\alpha = 1 - \beta$ (β being the discount factor) and p (period length) was derived analytically, visualized and commented. Attainable values of α_{HW} are limited by 0.5 from above. In comparison to DES as to a special case of Holt method, larger values of γ_{HW} are implied for the given value of α_{HW}. Moreover, very high values of δ_{HW}, strongly depending on p, are implied.

So the smoothing constant combinations attainable by the DLS approach are quite non-standard in light what typically gives the best results or is recommended as a routine fixed choice, see e.g. Chatfield and Yar (1988) or Cipra (2008), p. 306. Thus we do not expect this DLS approach to have large practical importance where the general Holt-Winters method is used. However, in some specific situations it could be used and then it is good to understand its relation to the general Holt-Winters method (which was the purpose of this chapter).

Chapter 6

Autocorrelated component in time series decomposition

6.1 Introduction

In this chapter we suggest extending the classical exponential smoothing type methods (SES, Holt and Holt-Winters method) by adding an *autocorrelated component* to their decomposition schemes. The autocorrelated component which causes a short-term autocorrelated variation around the trend can be considered as a compromise between long-term trend on one side and uncorrelated residual component on the other side.

By incorporating the autocorrelated component explicitly in the exponential smoothing method, we obtain a more complex method with some more parameters to be specified. In some situations we can gain an extra forecasting accuracy by doing so, preventing us from some common problems, e.g. from strongly positively autocorrelated forecasting errors. The importance of assuming an autocorrelated component could be higher in the case of data observed with higher frequency or in the case of irregularly observed data where the time step length is generally not limited from below.

In Section 6.2 the nature of an autocorrelated component is described in context of decomposition methods for smoothing and forecasting time series. Examples of real time series where the autocorrelated component could play an important role are mentioned. The impacts of ignoring presence of autocorrelated component and using classical methods are discussed.

In Section 6.3 the idea how to represent the autocorrelated component is introduced on the example of simple exponential smoothing. The relation of this extended method to Holt method with damped linear trend is noted. The version of the method for irregular time series is presented.

In Section 6.4 the way how to incorporate the autocorrelated component into general exponential smoothing type method is described. The possible choice between additive and multiplicative autocorrelated component is discussed.

Section 6.5 presents a numerical example of applyiing SES with and without autocorrelated component to a real time series (namely to the series of daily maximum temperatures in Sofia). Section 6.6 brings the conclusions of the chapter.

6.2 Autocorrelated component of time series

Decomposition time series methods. All exponential smoothing type methods are classical decomposition methods which means that they try to decompose the observed data into individual components with characteristic properties (usually trend, seasonal component and residual component). Based on a concrete decomposition scheme applied, smoothed values and forecasts are then constructed in an intuitive way.

There are various different decomposition schemes determined by presence or absence of individual components (e.g. seasonal or non-seasonal methods), by the way in which they are put together (e.g. additive or multiplicative decomposition) or just by the way in which individual components are modeled (e.g. linear or exponential trend; seasonality modeled by seasonal indices or trigonometric functions), see Section 2.1.

Trend and seasonal component are called *systematic* components and are used to construct smoothed values and forecasts. The rest of the observed data is called *residual (irregular, error, noise)* component and is treated as *non-systematic* part of the observed time series, usually assumed to be a realization of white noise process. This means specially that the values of residual component are not correlated in time.

Autocorrelated component as a semi-systematic one. There are some situations in reality when the residual component of the series is really uncorrelated in time. An example is the *measurement error* which can happen e.g. when the true value of a variable is estimated repeatedly over time via surveys using different samples from the whole concerned population. However, in most situations, beside the truly uncorrelated component (if such exists at all), there is an *autocorrelated component* which causes a short-term variation around the trend in a positively autocorrelated way. This variation has not a fixed period length as the seasonal one and also its pattern can change in time. We can alternatively call it a *cyclical* component.

A nice example is a time series of daily measured average temperature at a given place. There is a natural annual seasonality in such a series and the trend component can be thought as a constant determined by the local climate. When we subtract the seasonal component (which is very close to sinusoidal curve) we obtain "residuals" which are definitely not uncorrelated. We can observe a strongly positively autocorrelated cyclical variation caused by alternating of warm and cold waves usually after one or two weeks. The truly residual component plays not so important role here (if there is any), see Section 6.5.

In real numerical examples, while trying to construct forecasts of future unknown observations, it is sometimes difficult to distinguish the autocorrelated component from the trend component plus seasonal component on one side and the truly residual component on the other side. We can think about it as a *semi-systematic* component which is a compromise between the two ones mentioned above. While the uncorrelated residual component is "forgotten" after each new observation and so it has no time duration, the trend (plus seasonal) component are systematic which means that they last from now into future (of course with possible random innovations). The autocorrelated component preserves itself just in short-term horizons and is almost unpredictable concerning longer time horizons.

It is useful to admit that the adjectives *systematic* or *uncorrelated* are not absolute but are relative to the frequency in which the observations are taken. With a high observation frequency everything seams to be correlated and systematic and probably nearby nothing will remain as a true residual component. But with a low frequency some components of the series will appear to be uncorrelated in time, i.e. non-systematic.

This can be illustrated using an AR(1) process with the autoregressive parameter $\varphi \in (0,1)$. When we sample it regularly with time step m, we obtain an AR(1) process with parameter φ^m. Therefore when the observation frequency is low, the autocorrelation becomes negligible. On the other hand some real time series which has been created by regular sampling of some continuous-time variable can appear to be a realization of white noise process. However, if we increase the observation frequency, we will probably discover a significant autocorrelation in the observed values.

When we would use classical exponential smoothing methods which do not take the autocorrelated component into account we can face serious problems. There are two basic extreme ways how such a method can handle the autocorrelated component. It is either supposed to be a part of the residual component or it is added to the trend component. In the first case we will produce short-term forecasts with strongly autocorrelated forecasting errors while the long-term forecasts will be unaffected. In the second case the short-term forecasts are sufficiently accurate while the long-term forecasts are not. Of course that in reality we will probably face to some mixture of these two basic impacts.

6.3 Simple exponential smoothing

Simple exponential smoothing (SES) is the simplest exponential smoothing method. It assumes just a locally constant trend and a residual component. Only one parameter, smoothing constant α, has to be specified. Its optimal value reflects the ratio between the residual component variance and the variance of trend innovations. It can be approximately linked to the first order autocorrelation of the first differences of the series, see Hanzák (2007).

Inspiration by AR(1) process. Forecasts made at a fixed time for different time horizons form a horizontal straight line at the level of the latest smoothed value of the series. The difference between the last observation and this smoothed value is assumed to be a truly residual component with no effect for forecasts construction, not depending on the time horizon concerned. But it seams reasonable for the shorter horizon forecasts to be closer to the last observed value while longer horizon forecasts should tend to a certain limit.

It is known to forecasters that using a lower α value than optimal leads to a positive autocorrelation of forecasting errors and vice versa. Usually the value which is optimal in the sense of minimal MSE produces also almost uncorrelated forecasting errors. But this may not be the rule. If the series is generated by AR(1) process with parameter $\varphi \in (0, 1)$ then the MSE-optimal value of α and the value which would produce uncorrelated forecasts are different (concretely the second one is higher). So even if we use the MSE-optimal value of smoothing constant, our forecasting errors are positively correlated. This is a signal that the classical SES is not an optimal forecasting method for time series generated by stationary AR(1) processes.

In fact it is not surprising since the AR(1) process has different structure than one assumed by the method. There is a stable constant long-term trend, the centralized series can be thought as an autocorrelated component and there is no residual component at all. The optimal forecast of a future unknown observation $y_{t+\tau}$ made from time t is $\hat{y}_{t+\tau}(t) = \varphi^\tau y_t$. So the short-term forecasts are close to y_t which is also the smoothed value at time t while the forecasts tend to the (long-term) trend of 0 as $\tau \to \infty$.

Thus these forecasts have exactly the properties which those by SES lack of. Real non-seasonal time series with locally constant trend can be thought as a mixture of the ARIMA(0, 1, 1) model (for which SES is optimal) and the AR(1) process. They have a changing trend, autocorrelated cyclical component and uncorrelated truly residual component as well.

Model-based approach could consist in adding an AR(1) component to so called *basic structural model* (random walk plus noise) and applying the Kalman filter to construct smoothed values and forecasts in this extended state space model. But here we will follow the *ad hoc* approach typical for the original introduction of exponential smoothing methods.

Method formulation. Let us consider a univariate regular time series $\ldots, y_{t-1}, y_t, y_{t+1}, \ldots$. We assume the following decomposition scheme:

$$y_t = L_t + C_t + \varepsilon_t\,, \tag{6.1}$$

where L_t is the level at time t, C_t is the newly concerned *autocorrelated component* value at time t and ε_t is the residual component at time t. Besides the future random innovations, the trend L is constant in future, the autocorrelated component C decays exponentially to 0 with a *discount factor (damping constant)* $\varphi \in (0, 1)$ and the residual ε has no effect. This together implies the following forecasting formula:

$$\hat{y}_{t+\tau}(t) = L_t + \varphi^\tau C_t\,. \tag{6.2}$$

In particular, the smoothed value at time t is

$$\hat{y}_t = L_t + C_t \tag{6.3}$$

with the rest supposed to be a residual component:

$$\varepsilon_t = y_t - \hat{y}_t\,. \tag{6.4}$$

Now we will set up recurrent formulas for the update of individual components when moving from time t to time $t+1$ and after receiving the new observation y_{t+1}. We will follow the idea of error-correction formulas of classical exponential smoothing. Let us again denote

$$e_{t+1} = e(t)_{t+1} = y_{t+1} - \hat{y}_{t+1}(t) \tag{6.5}$$

the one-step-ahead forecasting error from time t. Let $\alpha \in (0, 1)$ be a smoothing constant for level and $\gamma \in [0, 1]$ a smoothing constant for the autocorrelated component. The level L is updated in the same way as in the classical method, see (2.2):

$$L_{t+1} = L_t + \alpha e_{t+1}\,. \tag{6.6}$$

The autocorrelated component C will be updated similarly as the seasonal component in Holt-Winters method, see (2.28), respecting its damping

$$C_{t+1} = \varphi C_t + \gamma(1-\alpha)e_{t+1}\,. \tag{6.7}$$

This guaranties us that the smoothed value $\hat{y}_{t+1}$ is always a convex linear combination of the forecast $\hat{y}_{t+1}(t)$ and the observed value y_t:

$$\hat{y}_{t+1} = (1-\gamma)(1-\alpha)\hat{y}_{t+1}(t) + [1-(1-\gamma)(1-\alpha)]\,y_{t+1}\,. \tag{6.8}$$

In case that this method is appropriate for the particular time series y, we can assume that the forecasting errors $\{e_t, t \in \mathbb{Z}\}$ form a white noise sequence. Then the level L exhibits a random walk, see (6.6), the autocorrelated component is driven by AR(1) process with parameter φ, see (6.7) and the residual component is a white noise

$$\varepsilon_t = (1-\gamma)(1-\alpha)e_t\,. \tag{6.9}$$

Based on the white noise assumption, we can easily derive the variance formula for forecasting errors with longer than unit horizon. Adding e.g. normality into our assumptions, the derivation of prediction intervals is straightforward.

The three parameters φ, α and γ are of course not known in reality and must be either chosen in an expert way or estimated from data, e.g. minimizing a certain forecasts accuracy criterion like MSE. When $\gamma = 0$ is taken, one gets the classical SES (if the initial value of C is set to 0).

The initialization of the method can be done using some simple approach similar to those used for classical method. Especially one can take L_0, the initial value of L, exactly like in the classical method (first observation, an average of several starting observations etc.) and initialize the autocorrelated component by taking $C_0 = 0$ or alternatively $C_0 = y_1 - L_0$ or $C_0 = \gamma(y_1 - L_0)$. More sophisticated initialization schemes are also possible. Another possible approach consists in backcasting.

Relation to Holt method with damped trend. There is an interesting relation between the suggested method and the Holt method with damped linear trend. Let us denote by L_t^H the level and by T_t the slope of Holt method at time time t and let α_H and γ_H be the corresponding smoothing constants. Then if we take

$$\alpha_H = \alpha + \gamma(1-\alpha)\,, \tag{6.10}$$

$$\gamma_H = \frac{\left(\frac{\alpha}{\alpha_H} - 1\right)(1-\varphi)}{\varphi} \tag{6.11}$$

and finally the slope damping constant of the Holt method being the same as the autocorrelated component damping constant φ, we obtain the same smoothing and forecasting results (for all horizons) as from the suggested method with parameters φ, α and γ. Further it holds

$$L_t^H = L_t + C_t \quad \text{and} \quad T_t^H = -\frac{1-\varphi}{\varphi} C_t\,. \tag{6.12}$$

Notice that $\alpha_H \in (0,1)$ but always $\gamma_H < 0$. Because of this fact and mainly because of the quite different interpretation, it is better to introduce them as two different methods rather than as two special cases of one method.

Version for irregular time series. Following the approach of Wright (1986), we can introduce the version of the suggested method for irregular time series. Let $\{y_{t_n}, n \in \mathbb{Z}\}$ be such a time series, $t_{n+1} > t_n$, $n \in \mathbb{Z}$. We will modify our updating formulas (6.6) and (6.7) into the form

$$L_{t_{n+1}} = L_{t_n} + \alpha_{t_{n+1}} e_{t_{n+1}} \tag{6.13}$$

and

$$C_{t_{n+1}} = \varphi^{t_{n+1}-t_n} C_{t_n} + \gamma_{t_{n+1}}(1-\alpha_{t_{n+1}}) e_{t_{n+1}}\,, \tag{6.14}$$

where

$$e_{t_{n+1}} = e(t_n)_{t_{n+1}} = y_{t_{n+1}} - \hat{y}_{t_{n+1}}(t_n) \tag{6.15}$$

is the forecasting error from time t_n to time t_{n+1} and the variable smoothing coefficients α_{t_n} and γ_{t_n} are updated in a recurrent way as in (2.15):

$$\alpha_{t_{n+1}} = \frac{\alpha_{t_n}}{\alpha_{t_n} + (1-\alpha)^{t_{n+1}-t_n}}\,, \tag{6.16}$$

$$\gamma_{t_{n+1}} = \frac{\gamma_{t_n}}{\gamma_{t_n} + (1-\gamma)^{t_{n+1}-t_n}}\,. \tag{6.17}$$

It is obvious that this generalizes the suggested method for regular time series and that it is also an extension of Wright's SES for irregular time series from Wright (1986) (see also Section 2.1).

6.4 General method

In this section we will show how to incorporate an autocorrelated component into any exponential smoothing type method. Moreover we will discuss the possibility of having more than one autocorrelated component in a decomposition or the possible multiplicative autocorrelated component. Finally the suggested method extension is compared with an existing approach how to cope with autocorrelated forecasting errors.

Let us consider any smoothing and forecasting method belonging to exponential smoothing family. We will consider its version for irregular time series $\{y_{t_n}, n \in \mathbb{Z}\}$. Let $\boldsymbol{S}_{t_n} = \left(S^{(1)}_{t_n}, S^{(2)}_{t_n}, \ldots, S^{(k)}_{t_n}\right)$ be a k-dimensional state vector of the method, e.g. $(L_{t_n}, T_{t_n}, \alpha_{t_n}, \delta_{t_n})$ in the case of Holt method. Let the original method be expressed by equations

$$\hat{y}_{t_n} = Y(\boldsymbol{S}_{t_n}), \quad (6.18)$$

$$\hat{y}_{t_n+\tau}(t_n) = F(\boldsymbol{S}_{t_n}, \tau), \quad (6.19)$$

$$e_{t_{n+1}} = y_{t_{n+1}} - \hat{y}_{t_{n+1}}(t_n), \quad (6.20)$$

$$\boldsymbol{S}_{t_{n+1}} = \boldsymbol{U}(\boldsymbol{S}_{t_n}, t_{n+1} - t_n, e_{t_{n+1}}), \quad (6.21)$$

where Y, F and $\boldsymbol{U}$ are given functions of appropriate dimensional arguments and values.

Now we add the autocorrelated component C to the current state vector $\boldsymbol{S}$ of the method and modify formulas (6.18) and (6.19) in the following way:

$$\hat{y}_{t_n} = Y(\boldsymbol{S}_{t_n}) + C_{t_n}, \quad (6.22)$$

$$\hat{y}_{t_n+\tau}(t_n) = F(\boldsymbol{S}_{t_n}, \tau) + \varphi^{\tau} C_{t_n}. \quad (6.23)$$

We must add an updating formula for C:

$$C_{t_{n+1}} = \varphi^{t_{n+1}-t_n} C_{t_n} + \gamma_{t_{n+1}} \left[y_{t_{n+1}} - Y(\boldsymbol{S}_{t_{n+1}})\right], \quad (6.24)$$

where γ_{t_n} is a smoothing coefficient updated by (6.17). Initialization of this extended method can be done analogously as in the case of SES in Section 6.3.

Multiple autocorrelated components. Similarly as it is possible and sometimes reasonable to have more than one seasonal component in Holt-Winters decomposition (see the daily and weekly seasonal component in Taylor (2003)), it is also possible to consider more than one autocorrelated component with different values of damping constant φ. These different values of φ refer to different average cycle lengths of the individual autocorrelated components (the higher φ is, the longer the average cycle length is). The way how to incorporate two or more autocorrelated components into a method is analogous to adding just one but the notation becomes more complicated when trying to put it down formally.

Multiplicative autocorrelated component. Similarly as we distinguish additive and multiplicative seasonal component, we can do so with autocorrelated component. Up to now we have worked with additive one where the amplitude of the autocorrelated component did not depend on the current level of the series, see (6.22)-(6.24). Probably the multiplicative autocorrelated component won't have so large practical importance when compared with seasonal one because the average cycle length is usually shorter, the pattern is not regular and the amplitude of an autocorrelated component is usually lower than that one of seasonal component.

However, in some specific situations it could be useful to let the autocorrelated component join the decomposition scheme in a multiplicative manner. In such a case we can change formulas (6.22)-(6.24) to

$$\hat{y}_{t_n} = Y(\boldsymbol{S}_{t_n}) \exp(C_{t_n}) \,, \tag{6.25}$$

$$\hat{y}_{t_n+\tau}(t_n) = F(\boldsymbol{S}_{t_n}, \tau) \exp\left(\varphi^{\tau} C_{t_n}\right) \,, \tag{6.26}$$

$$C_{t_{n+1}} = \varphi^{t_{n+1}-t_n} C_{t_n} + \gamma_{t_{n+1}} \{\ln(y_{t_{n+1}}) - \ln[Y(\boldsymbol{S}_{t_{n+1}})]\} \,, \tag{6.27}$$

see (4.28) and (4.29) for comparison. Of course it has sense only for positive data and only in the case when the logarithmic transformation was not used to convert the overall nature of the data from multiplicative to additive.

Comparison with an alternative approach. As we already mentioned, it often occurs that the one-step-ahead forecasting errors produced by the classical methods are positively correlated, see for example Taylor (2003). This can be thought as a result of autocorrelated component ignored by the method. It is obvious that these forecasts with autocorrelated errors are not optimal since there is an information which is systematically ignored by the method.

The classical way how to overcome these difficulties is as follows, see Chatfield and Yar (1988). We first apply the classical method to the time series. Then we take the one-step-ahead forecasting errors obtained doing so and calculate their first order sample autocorrelation coefficient, let us denote it by r. Then we always add the amount of $r^{\tau} e_t$ to the classical forecast for τ time units ahead from time t, where e_t is the last forecasting error occurred. The final forecasting errors will be uncorrelated and the accuracy of the method will be improved (new RMSE value should appr. equal $\sqrt{1 - r^2}$ - times the original one).

This procedure has an advantage of a simple "step by step" implementation. On the other hand, it has the following disadvantages when compared with using a method with autocorrelated component:

- There can be a confusion about what is the "forecast" in fact.
- It in fact implicitly assumes that there is no truly residual component in the series (as we would be restricted to have $\gamma = 1$ in our suggested method).
- It is not clear what is a smoothed value. Or the relations between observations, smoothed values and forecasts are not always too pleasant.
- Smoothing coefficients of the methods are optimized first and then the damping constant is "optimized" (estimated) with the smoothing constants already fixed. A simultaneous optimization can provide a better results.
- It is not clear how to use this approach in the case of an irregular time series.

6.5 Numerical example

As a numerical example, we will use the time series of daily maximum temperatures in °C in Sofia, Bulgaria, for period from 1st May to 30th September 2013 (5 complete months, 153 observations)[1].

Of course this time series contains an annual seasonality so a method for seasonal time series should be used in general. However, we use the series values for relatively short period (less then half of the period length) where the seasonality is not a dominant feature of the series development (the hottest month in average was July with 32.2 °C and the coldest was September with 24.4 °C), see Figure 6.2. So it is possible to consider the series as having locally constant trend and no seasonality. There is also an apparent autocorrelated component in the series, formed by short term variations of weather conditions (there are appr. 10 consecutive periods of warmer and colder days).

Both the classical SES and SES with autocorrelated component were implemented in author's application DMITS, see Chapter 7. Classical method (initialization using the first 10 observations, MSE-optimal value of smoothing

[1]http://freemeteo.com. Accessed on 16th February 2014.

constant) gives the following results. Smoothing constant α was optimized to $\alpha = 0.6064$, achieved RMSE value is 3.1277 and the first order sample autocorrelation coefficient of one-step-ahead forecasting errors is 0.073. It is obvious that the method absorbed the autocorrelated (cyclical) component into the time series level by using a relatively higher value of α. This guarantees the autocorrelation coefficient to be not significantly positive (but e.g. $\alpha = 0.25$ would bring the autocorrelation of 0.389).

SES with one additive autocorrelated component (initialization using again the first 10 observations, MSE-optimal values of smoothing constants α and γ and the damping constant φ) gives the following results. Optimal smoothing constants are $\alpha = 0.2340$ and $\gamma = 0.8564$ and $\varphi = 0.6270$, achieved RMSE value is 3.0379 (which is better than from the classical method) and the autocorrelation coefficient is 0.003, i.e. almost perfect 0.

The value of γ is close to 1 which means that there is just little residual variation in the series and majority of the de-trended series is attributed to the autocorrelated component. If we fix the value of γ to 1, we get optimal $\alpha = 0.2508$ and $\varphi = 0.5516$ and RMSE value of 3.0393, i.e. very similar results. In general, it could be possible to eliminate γ parameter by setting it to 1.

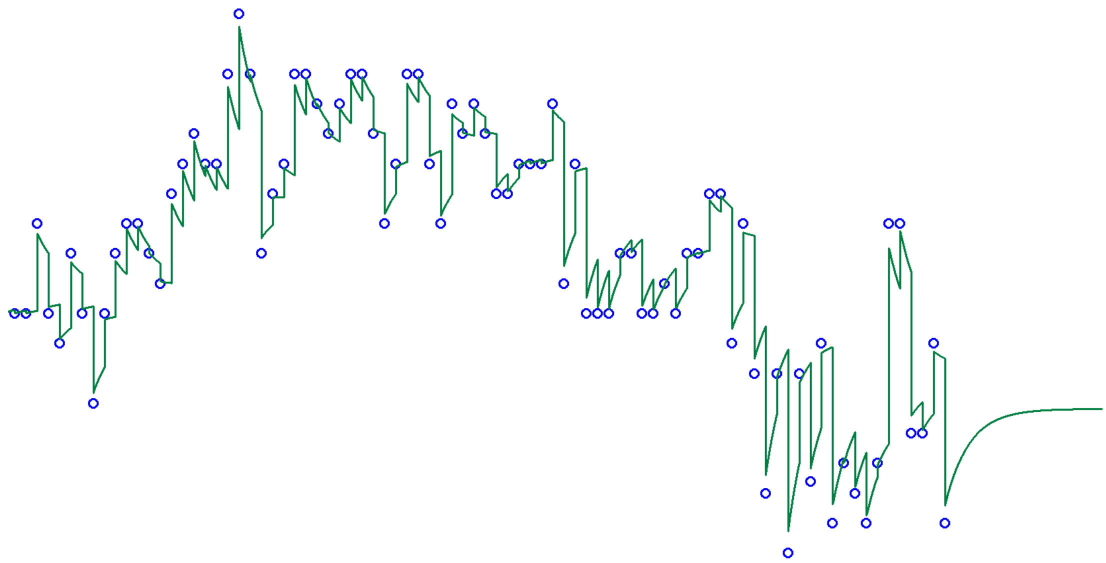

Figure 6.1: Daily maximum temperatures in Sofia, simple exponential smoothing with autocorrelated component (MSE-optimal smoothing constants).

Figure 6.1 shows the results of the method: actual observations (blue marks) and the smoothed and forecasted values (green line), zoomed for the period from July 9 till the end of the series. We can observe how the out-of-sample forecasts converge exponentially to the current level of the series, in contrast to the classical SES where such forecasts form a straight horizontal line.

Let us compare the values of the level smoothing constant α: 0.6064 vs. 0.2340. The classical method absorbed the changes in time series values into its trend (level) in much larger extend than the method with autocorrelated component.

On Figure 6.2 we can see how the method performed the series decomposition to the systematic component (trend), semi-systematic component (autocorrelated component) and residual component. We chose to plot decomposition as being forecasted from a previous observation point, i.e. we plotted values y_t (blue marks), $y_t(t-1) = L_{t-1} + \varphi C_{t-1}$ (trend + autocorrelated component; green line) and L_{t-1} (trend; red line) in the figure.

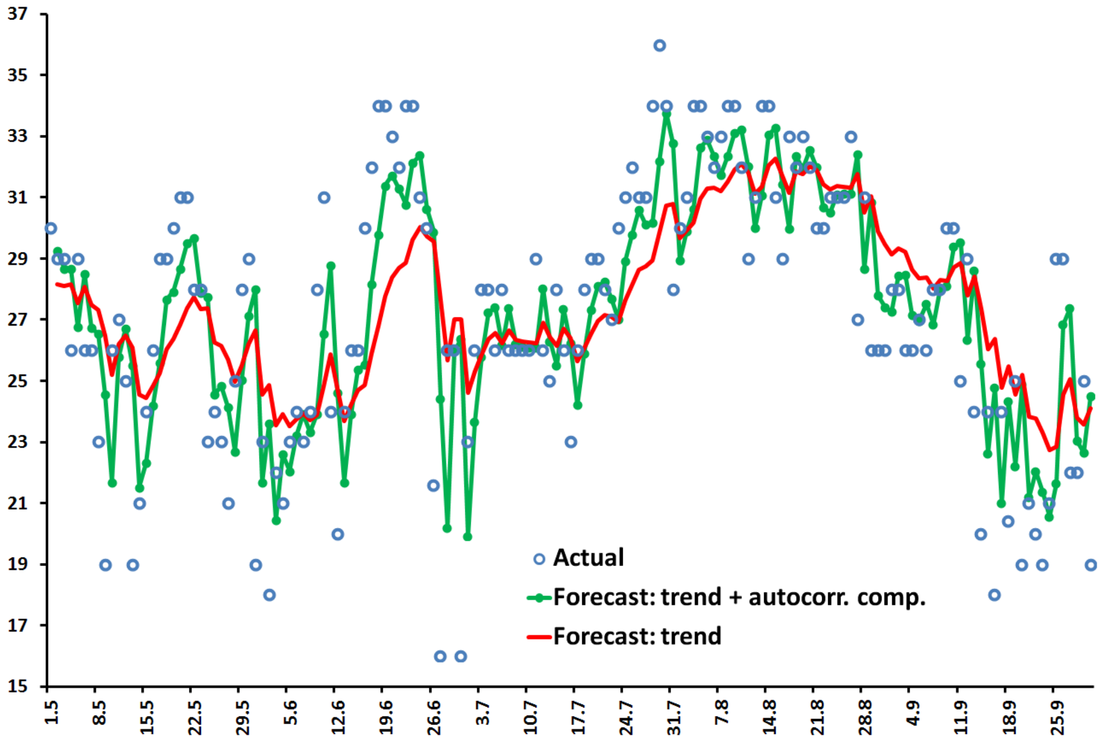

Figure 6.2: Daily maximum temperatures [°C] in Sofia, Bulgaria - time series decomposition using SES with autocorrelated component.

Naturally the red line is smoother than the green one, although it still absorbs part of the short term cyclical (autocorrelated) variation.

6.6 Conclusion

Autocorrelated component in time series decomposition can play the role of semi-systematic component, a hybrid between trend and residual component. It provides a solution to positively autocorrelated forecasting errors, being an integral part of the smoothing and forecasting method itself. We can also conclude that it brings an AR(1) building block into a time series decomposition scheme. It has some common features with seasonal component but the crucial difference is that it is not stable in long term, i.e. it does not have a fixed period but exhibits stochastic autocorrelated cycles of variable length (as AR(1) does). By setting $\gamma = 1$, the inclusion of autocorrelated component into the method just requires one additional parameter, namely the damping (or autocorrelation) parameter φ.

Chapter 7

Software implementation

One objective of the dissertation thesis was to transpose the suggested time series methods into a software form. To accomplish this objective, the application DMITS (**D**ecomposition **M**ethods for **I**rregular **T**ime **S**eries) was continuously developed as a part of this dissertation. It originated in the author's diploma thesis and now it includes other new methods suggested. The executable application is contained in the electronic attachment to this thesis, including a user manual in PDF.

The aim was to create user friendly and computationally robust software with various smoothing and forecasting methods available, all applicable to irregular time series. Detailed output is offered to the user, both in text and graphical form.

The software has been used in business practice for many years by me and my colleagues in MEDIARESEARCH research agency to smooth and forecast monthly, quarterly or yearly time series of continual surveys results.

Implementation of the time series methods involves the calculation of the initial values, the optimal choice of smoothing parameters, calculation of point and interval predictions and assessment of the accuracy and adequacy of the given predictive method used.

The methods available are:

- Simple exponential smoothing, see Section 2.1.
- Simple exponential smoothing with one additive autocorrelated component, see Section 6.3.
- Method based on an assumption of irregularly observed ARIMA$(0, 1, 1)$ process, see Section 2.3.
- Holt method with linear and damped linear trend, see Section 2.1.
- Modified Holt method with linear and damped linear trend, see Section 3.4.

- Double exponential smoothing, see Section 2.2.
- Triple exponential smoothing, see Section 2.2.
- DLS estimation of a linear trend, see Section 2.2.
- DLS estimation of a quadratic trend, see Section 2.2.
- Holt-Winters method with interpolated seasonal indices and trigonometric functions (additive or multiplicative seasonality), see Sections 4.2 and 4.3.

A fictitious observation time $t_0 = t_1 - q$ is always considered ($q = \frac{t_n - t_1}{n-1}$ is the average time spacing). The initial values are determined for this time point using DLS estimation of the trend in the initial section of the series, see Hanzák (2007) for details. Holt-Winters methods is an exception: its initial values are determined by back-casting. Prediction intervals are based on the assumption of normality of the forecasting errors, see Hanzák (2007) for details.

Now let us describe the general program functionalities and options, not depending on the particular method used. First we will describe what the user has to or may specify to the program, see also Figure 7.1. Later the output of the program will be described, see Figures 7.2 and 7.3.

Data input. As an input to the program, one must provide a sequence of time series observation values $y_{t_1}, y_{t_2}, \ldots, y_{t_n}$ and optionally a sequence of observation times $t_1, t_2, \ldots, t_n$. If the observation times are not specified, $t_j = j$ for all $j = 1, \ldots, n$ is used automatically as a default option (i.e. a regular time series is considered as a default). It must be $t_1 < t_2 < \ldots < t_n$, which is the only restriction on the observation times. There is a minimum number of observations (value n) required depending on the method used. A sequence of observation times $t_1, t_2, \ldots, t_n$ is not transformed by the program in any way so that any change in the time scale is the user's responsibility and must be made prior to entering the series into the program. Time series values $y_{t_1}, y_{t_2}, \ldots, y_{t_n}$ can generally be arbitrary real numbers unless a transformation with a limited domain is used (see below).

Transformation. Four different transformations can be optionally applied to the time series observations prior to the forecasting method is used. It is the logarithmic, square root, inverse logistic and inverse Gompertz transformation, see Hanzák (2007) for details. In the case of the first two transformations, all the observations must be positive. In the case of the remaining two transformations, all the observations must lie inside interval $(0, 1)$.

Choice of method. The choice of a particular smoothing and forecasting method is the user's responsibility. It has no effect on other options such as the usage of transformation etc. Of course, the methods differ in parameters that must be specified.

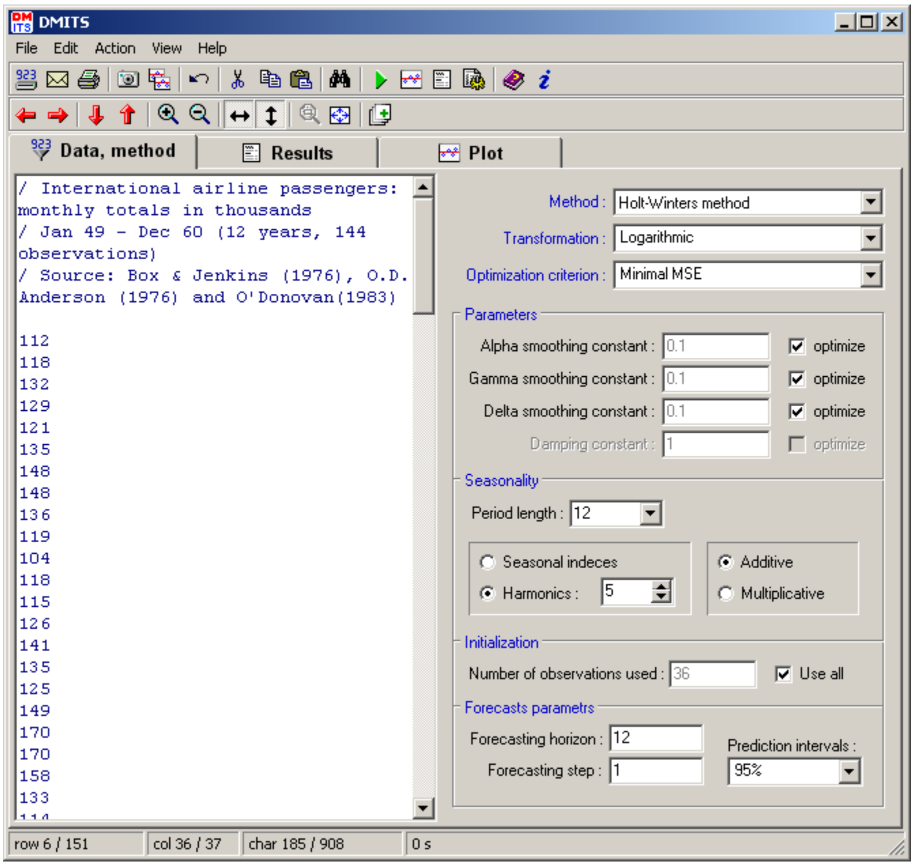

Figure 7.1: DMITS application: data input and method specification.

Choice of parameter values. Depending on the method used, one or three smoothing parameters must be specified. These are denoted as α, γ, δ and a *damping constant*. For each parameter, the user can either specify its fixed value or let the program find the optimal parameter value according to a given criteria. One, two or three parameters can be optimized simultaneously. By setting certain parameter to 0 or 1, a sub-method is obtained of the more general method (the damped linear trend changes to the classical linear one or the autocorrelated component in the simple exponential smoothing is omitted).

The program offers four possible optimization criteria: the minimum MSE (Mean Squares Error), MAE (Mean Absolute Error) and MAPE (Mean Absolute Percentage Error) and ML (maximum likelihood), see Hanzák (2007) for details. The optimization criterion is always calculated based on all n observations and their respective point predictions. To use MAPE criterion, all the observations must be positive.

Searching for the optimal combination of parameters in terms of the selected criterion is performed by an iterative numerical algorithm. Always all the parameters except one are set fixed and the criterion is optimized over the remaining parameter. The parameters being tuned rotate regularly. The optimization relies on the assumption of convexity of the minimized function. When this assumption is violated, it is possible that the solution will converge just to a local minimum which differs from the global minimum.

Number of observations for initialization. The user must specify the number of observations from the beginning of the series which will be used to determine the initial values for the recursive method. There is a minimum number of observations required depending on the method used. At most all the n observations can be used for the initialization. Initial values of the method depend on its parameters values so they are calculated repeatedly when the parameter(s) optimization is used.

Parameters of the forecasts. The length of the forecasting horizon and the time step for calculating the point predictions and prediction intervals must be specified. Forecasting horizon determines how far into future from the time of the last observation (t_n) the predictions will be calculated. These are plotted as a continuous curve into the time series plot and their values are also reported in the text output with the specified time step. The confidence level of the interval predictions can be chosen as 50, 75, 90, 95, or 99 %.

The text output of the program contains these sections:

Method specification. The first sections of the text output contains the information on the selected method. It is therefore largely a transcription of what the user has specified to be used. The only new information are the concrete values of the parameters that have been optimized.

Descriptive statistics of the time series. A basic information about the analyzed time series are reported: the number of observations n, the average time spacing q, mean, variance and standard deviation of the series observations.

Forecasting accuracy. The achieved values of several prediction accuracy measures are reported: MSE, RMSE (Root MSE), MAE and MAPE. They are calculated based on all n observations and their respective point forecasts. MAPE is calculated just in the case when all the observations are positive.

Forecasting method effectiveness. The prediction accuracy of the method used is compared to four trivial benchmark methods such as constant mean model or random walk model. The comparison is done based on the achieved MSE and it is expressed in terms of R-Squared, see Hanzák (2007) for details.

Forecasting method adequacy. It is tested whether the normalized prediction errors form a white noise. The mean and the mean square of these errors and the p-value for the zero mean test (one sample t-test) are reported. Further the first order sample autocorrelation and its significance p-value are reported, see Hanzák (2007) for details.

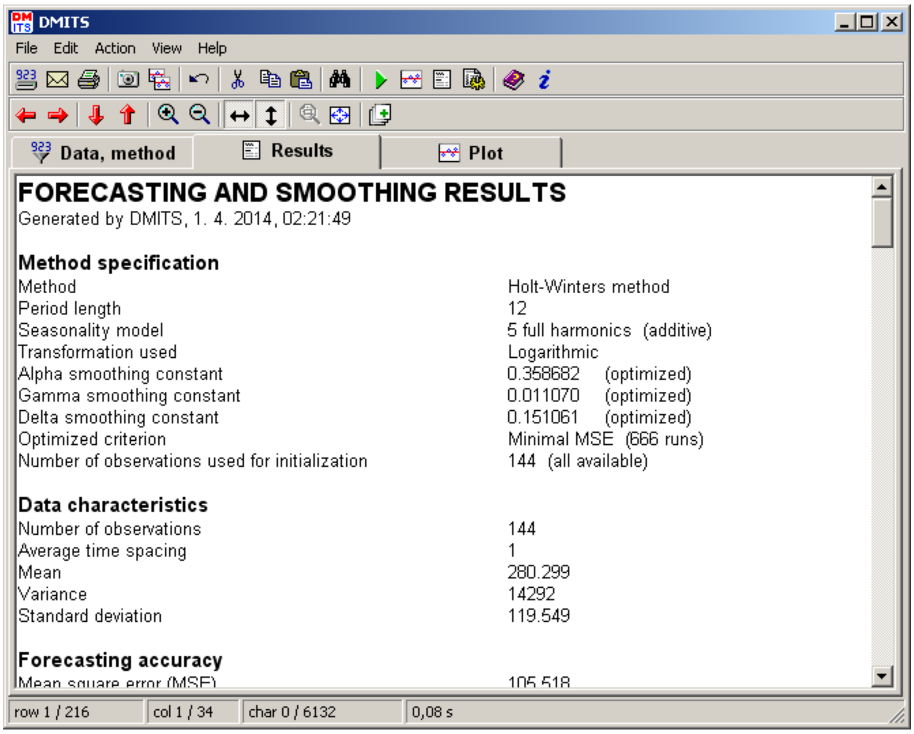

Figure 7.2: DMITS application: text output.

Forecasting errors normality tests. The sample skewness and kurtosis of the normalized forecasting errors and normality tests p-values based on these measures are reported, see Hanzák (2007) for details.

Prediction intervals performance. For all n observations it is evaluated whether they lie below, inside or above the respective prediction interval constructed from the previous observation time. These findings are then summarized in the form of three percentages compared with their theoretical counterparts (e.g. 5, 90 and 5 % for the 90% confidence level).

Historical forecasts and smoothed values. For $j = 1, \ldots, n$, the values j, t_j, y_{t_j}, $\hat{y}_{t_j}$ and $\hat{y}_{t_j}(t_{j-1})$ are reported.

Future forecasts. The point and interval predictions are reported for future times determined by the maximum forecasting horizon and the time step specified.

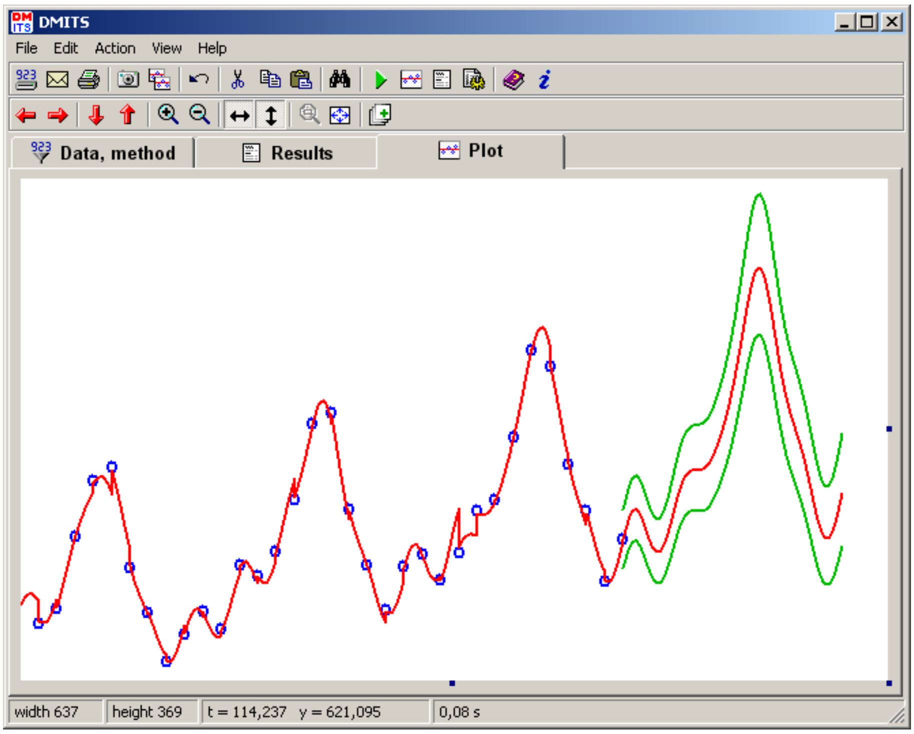

Figure 7.3: DMITS application: time series smoothing and forecasting plot.

Time series plot. The observation values (marks), historical smoothed values and forecasts (continuous curve) and future point predictions and interval predictions (three continuous curves) are plotted into a single graph. The historical smoothed values and forecasts are connected to form a continuous line in the following way. Inside the time interval $[t_{j-1}, t_j]$ the forecasts $\hat{y}_{t_{j-1}+\tau}(t_{j-1})$ for $\tau \in [0, t_j - t_{j-1}]$ are plotted. The two different values $\hat{y}_{t_j}(t_{j-1})$ and $\hat{y}_{t_j}$ that are plotted at time t_j are connected by a vertical line. This method of a graphical representation of historical forecasts and smoothed values seems to be more appropriate and useful than the usual linear interpolation of individual forecasts.

The plot margins adjust automatically to the plotted elements (however they can be fixed as well). Multiple plot results can be combined into one single compound plot to better visually compare different methods used. The plot can be shifted in horizontal or vertical direction, zoomed in one or both dimensions or zoomed into a selected rectangular area. It can be copied into the clipboard or saved as a graphical file.

Conclusions

Exponential smoothing type methods for smoothing and forecasting in time series are very popular in practice. The reason is their intuitive decomposition nature, adaptive performance and easy implementation through appropriate recursive formulas. When facing various types of irregularities in time series (missing observations, irregular time grid, outliers etc.) it is comfortable for the analysts that they can overcome these difficulties but still stay in the framework of the simple methods they are familiar with.

That is the reason why it has sense to make an effort in looking for suitable extensions of these classical (often *ad hoc*) methods even that there are more sophisticated model-based methods that could be used to do the job (e.g. based on state space models and Kalman filtering). Fortunately it appears that the empirical performance of the *ad hoc* and model-based methods is comparable.

In Chapter 3 a simple modification of Wright's version of Holt method for irregular time series (see Wright (1986)) was suggested. It has proved to be an efficient way to eliminate the unpleasant impact of time-close observations. The same simple modification of trend smoothing formula can be used in any special or extended version of Holt method, including its seasonal version, Holt-Winters method.

In Chapter 4 a general seasonality modeling concept was suggested in the framework of Holt-Winters method. Several particular seasonality settings were suggested, mainly interpolated seasonal indices and trigonometric functions. This can face not only the case of missing observations as Cipra et al. (1995) or Ratinger (1996) but also a generally irregular observation time grid. Moreover we receive broader possibilities to deal with seasonality even when working with regular time series. The usefulness of the suggested method were illustrated by real data examples and a simulation study.

In Chapter 5 it was shown that the DLS estimation of linear a regression with a linear trend and seasonal dummies is approximately equivalent to a certain special case of Holt-Winters method with additive seasonality. The dependence of the respective smoothing constants α_{HW}, γ_{HW} and δ_{HW} on the discount factor and period length was derived analytically and visualized. Maximum attainable values of α_{HW} and γ_{HW} are less than 1. Very high values of δ_{HW} are implied, not reflecting the usually more stable seasonal pattern of the series. Thus the practical importance of this approach stays rather limited.

In Chapter 6 an autocorrelated component as an additional term in the time series decomposition (beside the level, seasonal and residual component) was considered. This can solve the problem of positively autocorrelated one-step-ahead forecasting errors which often occurs even when using adaptive methods such as the exponential smoothing. The simple exponential smoothing with the autocorrelated component was illustrated using a real time series example.

Absence of a user friendly software available is in practice often a restrictive fact for application of a particular method. In Chapter 7 the author's software application DMITS, implementing most of the methods suggested in the thesis, was presented. The application is available to the reader in the electronic attachment to the thesis.

Practical time series analysts face to countless many data-specific challenges and sometimes no available method is capable to perform satisfactorily well in its original form. However, I believe that the methods suggested in this dissertation can solve some of the problems or at least serve as an inspiration to find a proper solution. Further development of the methods should ideally be motivated by practical experience with them and by new practical problems to solve.

Bibliography

Abraham, B. and Ledolter, J. (1983). *Statistical Methods for Forecasting.* Wiley, New York.

Aldrin, M. and Damsleth, E. (1989). Forecasting non-seasonal time series with missing observations. *Journal of Forecasting* **8**, 97–116.

Bessa, R. J. and Matos, M. A. (2011). Forecasting issues for managing a portfolio of electric vehicles under a smart grid paradigm. *2nd IEEE PES International Conference and Exhibition on "Innovative Smart Grid Technologies" (ISGT Europe 2011)*, 1–8.

Bessa, R. J. and Matos, M. A. (2012). Optimization Models for EV Aggregator Participation in a Manual Reserve Market. *IEEE Transactions on Power Systems* **28**, 3085–95.

Bessa, R. J. and Matos, M. A. (2014). Optimization models for an EV aggregator selling secondary reserve in the electricity market. *Electric Power Systems Research* **106**, 36–50.

Cipra, T. (1992). Robust exponential smoothing. *Journal of Forecasting* **11**, 57–69.

Cipra, T. (2006). Exponential smoothing for irregular data. *Applications of Mathematics* **51**, 597–604.

Cipra, T. (2008). *Financial Econometrics.* Ekopress, Prague (in Czech).

Cipra, T. and Hanzák, T. (2011). Exponential smoothing for time series with outliers. *Kybernetika* **47**, 165–178.

Cipra, T., Trujillo, J. and Rubio, A. (1995). Holt-Winters method with missing observations. *Management Science* **41**, 174–178.

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

Craig, J. C. (2004). B-splines. AE4375-6380 Supplemental Notes, Georgia Tech. Online at http://soliton.ae.gatech.edu/people/jcraig/classes/ae4375/notes/b-splines-04.pdf. Accessed on 9th February 2014.

Croux, C., Gelper, S. and Fried, R. (2008). Computational aspects of robust Holt-Winters smoothing based on M-estimation. *Applications of Mathematics* **53**, 163–176.

De Livera, A. M., Hyndman, R. J. and Snyder, R. D. (2011). A state space framework for automatic forecasting using exponential smoothing methods. *Journal of the American Statistical Association* **106**, 1513–1527.

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

Gardner, E. S. (2006). Exponential smoothing: The state of the art - Part II. *International Journal of Forecasting* **22**, 637–666.

Gardner, E. S. and McKenzie, E. (1985). Forecasting trends in time series. *Management Science* **31**, 1237–1246.

Gelper, S., Fried, R. and Croux, C. (2010). Robust forecasting with exponential and Holt-Winters smoothing. *Journal of Forecasting* **29**, 285–300.

Hanzák, T. (2007). *Decomposition methods for time series with irregular observations.* Diploma thesis, Faculty of Mathematics and Physics, Charles University in Prague. Supervisor Cipra, T. (in Czech).

Hanzák, T. (2008). Improved Holt method for irregular time series. *WDS'08 Proceedings of Contributed Papers, Part I - Mathematics and Computer Sciences*, Prague: Matfyzpress, 62—67.

Hanzák, T. (2012). Holt-Winters method with general seasonality. *Kybernetika* **48**, 1–15.

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

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting* **20**, 5–10.

Hyndman, R. J. (2010). Time series data library. Online at http://robhyndman.info/TSDL. Accessed on 26th June 2010.

Hyndman, R. J., Koehler, A. B., Snyder, R. D. and Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting* **18**, 439—454.

Mikulka, J. (2008). *Exponential smoothing.* Bachelor thesis, Faculty of Mathematics and Physics, Charles University in Prague. Supervisor Hanzák, T. (in Czech).

Qing Ye, Szeto, W. Y. and Wong, S. C. (2010). Short-Term Traffic Speed Forecasting Based on Data Recorded at Irregular Intervals. *EEE Transactions on Intelligent Transportation Systems* **13**, 1727–37.

Ratinger, T. (1996). Seasonal time series with missing observations. *Applications of Mathematics* **41**, 41–55.

Ruckdeschel, P., Spangl, B. and Pupashenko, D. (2014). Robust Kalman tracking and smoothing with propagating and non-propagating outliers. *Statistical Papers* **55**, 93-123.

Taylor, J. W. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of Operational Research Society* **54**, 799–805.

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

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science* **6**, 324–342.

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

List of author's publications

Hanzák, T. (2008). Improved Holt method for irregular time series. *WDS'08 Proceedings of Contributed Papers, Part I - Mathematics and Computer Sciences*, Prague: Matfyzpress, 62—67.

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

Cipra, T. and Hanzák, T. (2011). Exponential smoothing for time series with outliers. *Kybernetika* **47**, 165–178.

Hanzák, T. (2012). Holt-Winters method with general seasonality. *Kybernetika* **48**, 1–15.

List of Tables

List of Figures