

# MPRA

Munich Personal RePEc Archive

## **Complex Exponential Smoothing**

Svetunkov, Ivan and Kourentzes, Nikolaos

Lancaster University Management School

1 May 2015

Online at <https://mpra.ub.uni-muenchen.de/69394/>

MPRA Paper No. 69394, posted 15 Feb 2016 16:50 UTC

# Complex Exponential Smoothing

Ivan Svetunkov \*

and

Nikolaos Kourentzes

Department of Management Science, Lancaster University

August 28, 2015

## Abstract

Exponential smoothing has been one of the most popular forecasting methods for business and industry. Its simplicity and transparency have made it very attractive. Nonetheless, modelling and identifying trends has been met with mixed success, resulting in the development of various modifications of trend models. With the aim of addressing this problem, we propose the “Complex exponential smoothing” (CES), based on the new notion of “information potential”, an unobserved time series component, and the theory of functions of complex variables. The underlying statistical model of CES is described here. It has several advantages over conventional exponential smoothing models: it can model and forecast both stationary and nonstationary process; hence CES can capture both level and trend cases in the conventional exponential smoothing classification. This effectively sidesteps model selection for CES. It is evaluated on real data demonstrating better performance than established benchmarks and other exponential smoothing methods.

*Keywords:* Time Series, Forecasting, State Space Models, Exponential Smoothing, Complex variables

---

\*Correspondance: i.svetunkov@lancaster.ac.uk, Department of Management Science, Lancaster University Management School, Lancaster, Lancashire, LA1 4YX, UK.

# 1 Introduction

Exponential smoothing is a very successful group of forecasting methods which is widely used both in theoretical research (for examples see Jose and Winkler, 2008; Kolassa, 2011; Maia and de A.T. de Carvalho, 2011; Wang et al., 2012; Athanasopoulos and de Silva, 2012; Kourentzes et al., 2014) and in practise (see different forecasting comparisons by Fildes et al., 1998; Makridakis and Hibon, 2000; Gardner and Diaz-Saiz, 2008; Athanasopoulos et al., 2011).

The exponential smoothing methods are well known and popular amongst practising forecasters for more than half a century originating from the work by Brown (1956). Hyndman et al. (2002), based on work by Ord et al. (1997), embedded exponential smoothing within a state space framework, providing its statistical rationale, resulting in ETS (short for ExponentTial Smoothing). ETS provides a systematic framework to estimate parameters, construct prediction intervals and choose between different types of exponential smoothing.

While ETS is widely used, recent work has demonstrated that it is possible to improve upon. Research shows that various combinations of exponential smoothing models result in composite ETS forms beyond the 30 standard forms (Hyndman et al., 2008) which leads to improvement in forecasting accuracy (Kolassa, 2011; Kourentzes et al., 2014). We argue that models in the existing taxonomy do not cover all the possible forms. A possible complicating factor for ETS is the assumption that any time series can be decomposed into several distinct components, namely level, trend and seasonality. In particular separating the level and trend of the time series presents several difficulties. The motivation in this paper is to try to avoid this artificial distinction. For this reason we focus our research on the level and trend aspects, and do not examine seasonality that can usually be easily

distinguished.

We propose a new approach to time series modelling that eliminates the arbitrary distinction between level and trend components, and effectively sidesteps the model selection procedure. To do this we introduce the concept of “information potential”, an unobserved time series component. We argue that this information exists in any time series and including it in a model results in superior forecasting accuracy. We encode the observed values of a time series together with the information potential as complex variables, giving rise to the proposed Complex Exponential Smoothing (CES). We demonstrate that CES has several desirable properties in terms of modelling flexibility over conventional exponential smoothing and demonstrate its superior forecasting accuracy empirically.

The rest of the article is structured as follows. The conventional and the new approaches to time series modelling are briefly discussed in the section 2, leading to the introduction of the Complex Exponential Smoothing (CES) in the section 3. Within that section properties of CES and connections with existing models are discussed. Section 4 proposes an approximation for the information potential and discusses its implications for CES. Finally real life examples are shown and the comparison of CES performance with some other statistical models is carried out in the section 5, followed by concluding remarks.

## **2 Conventional and complex-valued approaches**

Hyndman et al. (2008) systematised all existing exponential smoothing methods and showed that any ETS model is based on one out of five types of trends (none, additive, damped additive, multiplicative and damped multiplicative), one out of two types of errors (additive and multiplicative) and one out of three types of seasonal components (none, additive and multiplicative). This taxonomy leads to 30 exponential smoothing models that underlay

different types of time series. Model parameters are optimised using maximum likelihood estimation and analytical variance expressions are available for most exponential smoothing forms. The authors proposed to use information criteria for model selection, which allowed choosing the most appropriate exponential smoothing model in each case. Hyndman et al. (2008) argued that AIC is the most appropriate information criterion. Billah et al. (2006) demonstrated that there was no significant difference in the forecasting accuracy when using different information criteria.

Focusing on the non-seasonal cases the general ETS data generating process has the following form:

$$y_t = f(l_{t-1}, b_{t-1}, \epsilon_t), \quad (1)$$

where  $y_t$  is the value of the series,  $l_t$  is the level component,  $b_t$  is the trend component,  $\epsilon_t$  is the error term and  $f(C(\cdot))$  is some function that allows including these components in either additive or multiplicative form. However the composite forecasts discussed in Kolassa (2011) and Kourentzes et al. (2014) hint to the lack of a clear separation between level and trend components. This decomposition in (1) can be considered arbitrary: depending on the chosen ETS model, its initial values and smoothing parameters, different estimates of the time series components can be obtained. Besides, these components are unobservable and their identification relies on an appropriate decomposition. For example it is often hard to distinguish a local-level time series from a trend series. Consider simple exponential smoothing (SES):

$$\hat{y}_t = \alpha y_{t-1} + (1 - \alpha)\hat{y}_{t-1}, \quad (2)$$

where  $\hat{y}_t$  is the estimated value of the series and  $\alpha$  is the smoothing parameter which should lie inside the region  $(0, 2)$  (Brenner et al., 1968). Following the notation by Hyndman et al.

(2002) SES has an underlying ETS(A,N,N) model:

$$\begin{cases} y_t = l_{t-1} + \epsilon_t \\ l_t = l_{t-1} + \alpha\epsilon_t \end{cases} \quad (3)$$

When the smoothing parameter in (3) increases, changes in level become so rapid that the generated time series can reveal features of trends. When the smoothing parameter becomes greater than one, the series can reveal an even clearer global trend (Hyndman et al., 2008, p.42). Selecting the correct model in this situation becomes a challenging task. This is just one example where it is hard to identify the correct model with the standard exponential smoothing decomposition approach.

Instead of decomposing time series we propose to study two characteristics: i) the observed value of the series  $y_t$ ; and ii) its “information potential”  $p_t$ . We introduce the information potential as a non-observable component of the time series that characterises it and influences the observed values  $y_t$ . We propose that although the actual value  $y_t$  and its prediction are the variables of main interest in modelling, the unobserved information potential  $p_t$  contains additional useful information about the observed series. Thus it is necessary to include it in models. In this paper we show that the information potential increases the flexibility of models allowing to capture a wider range of behaviours, removes the need for the arbitrary distinction between level and trend and results in increased forecasting accuracy.

We encode these two real variables into a single complex variable  $y_t + ip_t$  that permits both of them to be taken into account during the modelling process (Svetunkov, 2012). Here  $i$  is the imaginary unit which satisfies the equation:  $i^2 = -1$ . Thus the general data generating process using this notation has the form:

$$y_t + ip_t = f(Q, \epsilon_t), \quad (4)$$

where  $Q$  is a set of some complex variables, chosen for the prediction of  $y_t + ip_t$ . Using this idea, we propose the Complex Exponential Smoothing (CES) based on the data generating process (4), in analogy to the conventional exponential smoothing model. This is introduced in the detail in the following section.

### 3 Complex Exponential Smoothing

#### 3.1 Method and model

Combining the simple exponential smoothing method with the idea of information potential, substituting the real variables in (2) by complex variables, gives us the complex exponential smoothing method:

$$\hat{y}_{t+1} + i\hat{p}_{t+1} = (\alpha_0 + i\alpha_1)(y_t + ip_t) + (1 - \alpha_0 + i - i\alpha_1)(\hat{y}_t + i\hat{p}_t), \quad (5)$$

where  $\hat{y}_t$  is the estimated value of series,  $\hat{p}_t$  is the estimated value of information potential and  $\alpha_0 + i\alpha_1$  is complex smoothing parameter. Representing the complex-valued function as a system of two real-valued functions leads to:

$$\begin{cases} \hat{y}_{t+1} = (\alpha_0 y_t + (1 - \alpha_0)\hat{y}_t) - (\alpha_1 p_t + (1 - \alpha_1)\hat{p}_t) \\ \hat{p}_{t+1} = (\alpha_1 y_t + (1 - \alpha_1)\hat{y}_t) + (\alpha_0 p_t + (1 - \alpha_0)\hat{p}_t) \end{cases}. \quad (6)$$

It can be seen from (6) that the final CES forecast consists of two parts: one is produced by SES, while the second only employs the SES mechanism. It is obvious that CES is a non-linear method in its nature as both first and second equations in (6) are connected with each other and change simultaneously depending on the complex smoothing parameter value.

We derive the underlying statistical model for CES to study its properties. The model can be written in the following state-space form (see A for the derivation):

$$\begin{cases} y_t = l_{t-1} + \epsilon_t \\ l_t = l_{t-1} - (1 - \alpha_1)c_{t-1} - \alpha_1 p_t + \alpha_0 \epsilon_t , \\ c_t = l_{t-1} + (1 - \alpha_0)c_{t-1} + \alpha_0 p_t + \alpha_1 \epsilon_t \end{cases} \quad (7)$$

where  $l_t$  is the level component,  $c_t$  is the information component at observation  $t$  and  $\epsilon_t \sim N(0, \sigma^2)$ . Observe that dependencies in time series have a non-linear structure and no explicit trend component is present in the time series as this model does not need to artificially break the series into level and trend, as ETS does. This idea still allows to rewrite (7) in a shorter more generic way, resembling the general SSOE state-space framework:

$$\begin{cases} y_t = w' x_{t-1} + \epsilon_t \\ x_t = F x_{t-1} + q p_t + g \epsilon_t \end{cases} , \quad (8)$$

where  $x_t = \begin{pmatrix} l_t \\ c_t \end{pmatrix}$  is the state vector,  $F = \begin{pmatrix} 1 & -(1 - \alpha_1) \\ 1 & 1 - \alpha_0 \end{pmatrix}$  is the transition matrix,  $g = \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix}$  is the persistence vector,  $q = \begin{pmatrix} -\alpha_1 \\ \alpha_0 \end{pmatrix}$  is the information potential persistence vector and  $w = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$  is the measurement vector.

The state-space form (8) permits extending the CES in a similar ways to ETS to include additional states for seasonality or exogenous variables. The main difference between model (8) and the conventional ETS is that the former includes the information potential term. Furthermore the transition matrix in (8) includes smoothing parameters which is not a standard feature for ETS models. This form uses the generic SSOE state-space exponential smoothing notation and allows a clear understanding of the model.



## 3.2 Properties of CES

Being an exponential smoothing method CES is capable in distributing weight between the observations in a different manner. In this paragraph we will show how the complex smoothing parameter influences this distribution.

Substituting the estimated values in the right side of equation (5) repeatedly the following recursive form of CES can be obtained:

$$\hat{y}_{t+1} + i\hat{p}_{t+1} = (\alpha_0 + i\alpha_1) \sum_{j=0}^T (1 - \alpha_0 + i - i\alpha_1)^j (y_{t-j} + ip_{t-j}), \quad (9)$$

where  $T \rightarrow \infty$ .

To obtain the bounds of the complex smoothing parameter the weights in (9) can be perceived as a geometric progression with a complex ratio  $(1 - \alpha_0 + i - i\alpha_1)$ ; hence this series will converge to some complex number only if the absolute value of the ratio is less than one:

$$v = \sqrt{(1 - \alpha_0)^2 + (1 - \alpha_1)^2} < 1. \quad (10)$$

The condition (10) is crucial for the preservation of the stability condition (Hyndman et al., 2008): the older observations should have smaller weights compared to the newer observations. For CES this is represented graphically on the plane as a circle with a centre at coordinates  $(1, 1)$  as in Figure 1.

Using (9) CES can also be represented in the trigonometric form of complex variables, which should help in understanding of the mechanism used in the method:

$$\hat{y}_{t+1} + i\hat{p}_{t+1} = R \sum_{j=0}^T [v^j (\cos(\varphi + j\gamma) + i \sin(\varphi + j\gamma)) (y_{t-j} + ip_{t-j})], \quad (11)$$

where  $R = \sqrt{\alpha_0^2 + \alpha_1^2}$ ,  $\varphi = \arctan \frac{\alpha_1}{\alpha_0} + 2\pi k$  and  $\gamma = \arctan \frac{1-\alpha_1}{1-\alpha_0} + 2\pi k$ ,  $k \in Z$ . Hereafter we assume that  $k = 0$  in the calculation of all the polar angles as all other angles do not

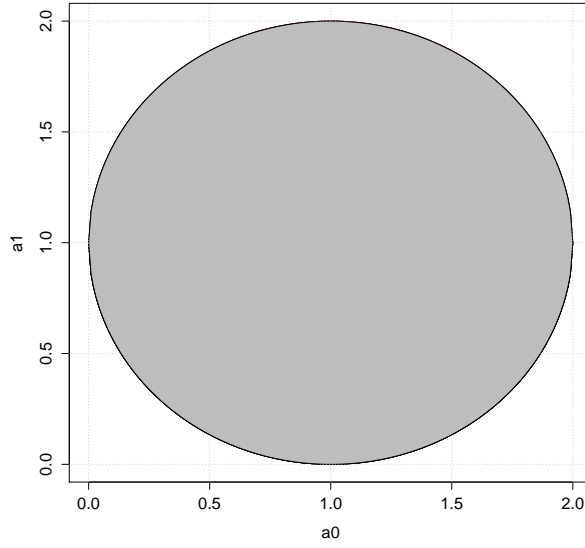


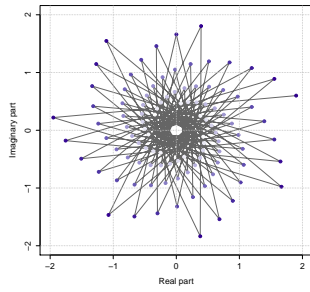
Figure 1: Bounds of complex smoothing parameter.

add to model understanding.

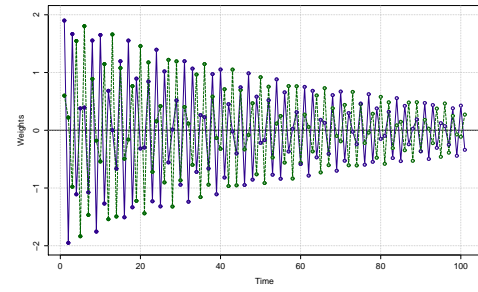
Multiplying complex variables in (11) shows how the real and imaginary parts of the forecast are formed separately, which can be presented in the following system:

$$\begin{cases} \hat{y}_{t+1} = R \sum_{j=0}^T v^j \cos(\varphi + j\gamma) y_{t-j} - R \sum_{j=0}^T v^j \sin(\varphi + j\gamma) p_{t-j} \\ \hat{p}_{t+1} = R \sum_{j=0}^T v^j \sin(\varphi + j\gamma) y_{t-j} + R \sum_{j=0}^T v^j \cos(\varphi + j\gamma) p_{t-j} \end{cases} \quad (12)$$

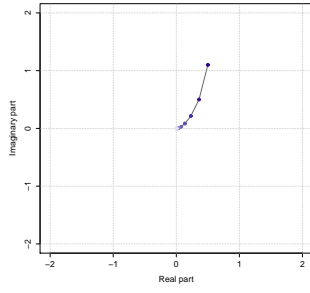
Here the CES forecast depends on the previous values of the series and the information potential, which are weighed in time using trigonometric functions, depending on the value of the complex smoothing parameter. For example, when  $\alpha_0 + i\alpha_1 = 1.9 + 0.6i$  weights are distributed as shown on Figures 2a and 2b: they oscillate and converge to zero slowly and their absolute value decreases in time due to (10). This weights distribution results in forecasts based on the long term time series characteristics. A more conventional case is



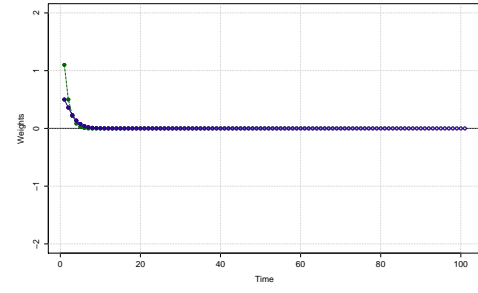
(a) on complex plane



(b) in time



(c) on complex plane



(d) in time

Figure 2: CES weights distribution. 2b and 2d: blue solid lines – real parts of complex weights, green dashed lines – the imaginary parts of complex weights.

shown in Figures 2c and 2d. The parameter is equal to  $0.5 + 1.1i$  which results in rapid exponential decline of the weights in time, using short-term characteristics of time series. Plots 2a and 2b show the decrease of weights on complex plane. They demonstrate what is happening with the equivalent complex smoothing parameter when it is exponentiated in the power  $j$  with  $j = 1, \dots, T$ : the original vector is rotated on the complex plane.

Plots on the Figure 2 also demonstrate that due to the complex nature of the weights their sum will usually be a complex number that can be calculated using:

$$S = (\alpha_0 + i\alpha_1) \sum_{j=0}^{\infty} (1 - \alpha_0 + i - i\alpha_1)^j = \frac{\alpha_0^2 - \alpha_1 + \alpha_1^2 + i\alpha_0}{\alpha_0^2 + (1 - \alpha_1)^2}. \quad (13)$$

The sum of weights depends solely on the value of the complex smoothing parameter. It can become real number only when  $\alpha_0 = 0$  which contradicts condition (10). It can also be concluded that the real part of  $S$  can be any real number while its imaginary part is restricted to positive real numbers only.  $S$  indicates that CES is not an averaging model, diverging from the conventional exponential smoothing interpretation.

### 3.3 Connection with other forecasting models

#### 3.3.1 Underlying ARIMAX model

The majority of exponential smoothing models have equivalent underlying ARIMA models. For example, ETS(A,N,N) has underlying ARIMA(0,1,1) model (Gardner, 1985). It can be shown that CES has an underlying ARIMAX(2,0,2) model (see Appendix B for the derivations):

$$\begin{cases} (1 - \phi_1 B - \phi_2 B^2)y_t = (1 - \theta_{1,1} B - \theta_{1,2} B^2)\epsilon_t + (-\gamma_1 B - \gamma_2 B^2)p_t, \\ (1 - \phi_1 B - \phi_2 B^2)\xi_t = (1 - \theta_{2,1} B - \theta_{2,2} B^2)p_t + (-\gamma_1 B - \gamma_2 B^2)\epsilon_t \end{cases}, \quad (14)$$

where  $\xi_t = p_t - c_{t-1}$  is an information gap, the value showing the amount of the information missing in the information component  $c_t$ , while the parameters of ARIMAX are:  $\phi_1 = 2 - \alpha_0$ ,  $\phi_2 = \alpha_0 + \alpha_1 - 2$ ,  $\theta_{1,1} = 2 - 2\alpha_0$ ,  $\theta_{1,2} = 2\alpha_0 + 2\alpha_1 - 2 - \alpha_0^2 - \alpha_1^2$ ,  $\theta_{2,1} = -2$ ,  $\theta_{2,2} = 2$ ,  $\gamma_1 = \alpha_1$ ,  $\gamma_2 = \alpha_0 - \alpha_1$ .

The first equation in (14) demonstrates how the actual data is generated, while the second equation shows how the unobservable part of the data is formed. The information potential variable  $p_t$  in the right hand side of the first equation of the model (14) functions as an external variable that contributes to the model and refines the AR and MA terms. The error term  $\epsilon_t$  in the second equation in (14) has the same role and coefficients as the information potential in the first equation. Furthermore the AR term has the same coefficients in both first and second equations in (14).

### 3.3.2 Comparison with single exponential smoothing

When  $p_t = 0$  and  $\alpha_1 = 1$  the system (6) becomes:

$$\begin{cases} \hat{y}_{t+1} = \alpha_0 y_t + (1 - \alpha_0) \hat{y}_t \\ \hat{p}_{t+1} = y_t + (1 - \alpha_0) \hat{p}_t \end{cases} \quad (15)$$

It is obvious that the first equation in (15) is the SES method and that the two equations in (15) become independent from each other under these conditions. The second equation is not needed for the purpose of forecasting.

The interpretation of  $p_t = 0$  is that no additional information is used in the time series generation and the classical exponential smoothing time series decomposition view

is adequate. This results from (7) in the following CES model (see appendix C):

$$\begin{cases} y_t = l_{t-1} + \epsilon_t \\ l_t = l_{t-1} + \alpha_0 \epsilon_t \\ c_t = \frac{l_{t-1}}{\alpha_0} + \frac{\epsilon_t}{\alpha_0} \end{cases} \quad (16)$$

We see in (16) that the level component becomes the same as in ETS(A,N,N). The only difference of CES is the presence of non-observable information component that does not interact with  $y_t$ .

In this case the weights in CES are distributed similarly to single exponential smoothing and converge to the complex number (13) which becomes equal to:

$$S = \frac{\alpha_0^2 + i\alpha_0}{\alpha_0^2} = 1 + i\frac{1}{\alpha_0}. \quad (17)$$

The real part of  $S$  is equal to 1, as in SES. The stability region of CES becomes equivalent to the stability region of SES, a (0, 2) region for the smoothing parameter:

$$\begin{aligned} v &= \sqrt{(1 - \alpha_0)^2 + (1 - \alpha_1)^2} < 1, \\ v &= |1 - \alpha_0| < 1, \text{ with } \alpha_1 = 0. \end{aligned} \quad (18)$$

From the above we see that CES encompasses SES and that it is the information potential part that gives CES its different properties.

## 4 Information potential approximation

### 4.1 State-space form

Due to the unobservability of the information potential it needs to be approximated by some characteristics of the studied time series. An approximation that leads to several

useful properties is:

$$p_t = y_t - \hat{y}_t = \epsilon_t. \quad (19)$$

The logic behind this approximation is that  $\epsilon_t$  can be seen as a gauge of the information not included in a model.

This assumption leads to a different state-space model, based on (7), underlying CES:

$$\begin{cases} y_t = l_{t-1} + \epsilon_t \\ l_t = l_{t-1} - (1 - \alpha_1)c_{t-1} + (\alpha_0 - \alpha_1)\epsilon_t \\ c_t = l_{t-1} + (1 - \alpha_0)c_{t-1} + (\alpha_0 + \alpha_1)\epsilon_t \end{cases} \quad (20)$$

Now the persistence vector  $g = \begin{pmatrix} \alpha_0 - \alpha_1 \\ \alpha_0 + \alpha_1 \end{pmatrix}$ , while all the other vectors and matrices values from (7) can be retained. The main difference between the general state-space CES (7) and (20) is that the smoothing parameter changes in both equations of the transition equation, leading to additional non-linearity in the model. The conditional variance of the model can be estimated using the new persistence vector and the the transition matrix can be used to estimate the conditional mean.

## 4.2 Likelihood function

The error term in (20) is additive. As a result the likelihood function for CES is trivial and is similar to the likelihood function of additive exponential smoothing models (Hyndman et al., 2008, p.68):

$$L(g, x_0, \sigma^2 | y) = \left( \frac{1}{\sigma\sqrt{2\pi}} \right)^T \exp \left( -\frac{1}{2} \sum_{t=1}^T \left( \frac{\epsilon_t}{\sigma} \right)^2 \right). \quad (21)$$

It was shown by Hyndman et al. (2008) that when the error term is distributed normally, maximizing the likelihood function to estimate the parameters of CES is equivalent to

minimizing the sum of squared errors:  $SSE = \sum_{t=1}^t \epsilon_t^2$ .

### 4.3 Stationarity and stability conditions for CES

The stationarity condition for general exponential smoothing in the state-space form (8) holds when all the eigenvalues of  $F$  lie inside the unit circle (Hyndman et al., 2008, p.38). CES can be both stationary and not, depending on the complex smoothing parameter value, in contrast to ETS models that are always nonstationary. Calculating eigenvalues of  $F$  for CES gives the following roots:

$$\lambda = \frac{2 - \alpha_0 \pm \sqrt{\alpha_0^2 + 4\alpha_1 - 4}}{2}. \quad (22)$$

If the absolute values of both roots are less than 1 then the estimated CES is stationary.

When  $\alpha_1 > 1$  one of the eigenvalues will always be greater than one. In this case both eigenvalues will be real numbers and CES produces a non-stationary trajectory. When  $\alpha_1 = 1$  CES becomes equivalent to ETS(A,N,N). Finally, the model becomes stationary when (see appendix D):

$$\begin{cases} \alpha_1 < 5 - 2\alpha_0 \\ \alpha_1 < 1 \\ \alpha_1 > 1 - \alpha_0 \end{cases} \quad (23)$$

The other important property that arises from (20) is the stability condition for CES. We first use  $\epsilon_t = y_t - l_{t-1}$  in the transition equation in (20). After manipulations, the following is obtained:

$$\begin{cases} y_t = l_{t-1} + \epsilon_t \\ \begin{pmatrix} l_t \\ c_t \end{pmatrix} = \begin{pmatrix} 1 - \alpha_0 + \alpha_1 & -(1 - \alpha_1) \\ 1 - \alpha_0 - \alpha_1 & 1 - \alpha_0 \end{pmatrix} \begin{pmatrix} l_{t-1} \\ c_{t-1} \end{pmatrix} + \begin{pmatrix} \alpha_0 - \alpha_1 \\ \alpha_1 + \alpha_0 \end{pmatrix} y_t \end{cases}. \quad (24)$$



The matrix  $D = \begin{pmatrix} 1 - \alpha_0 + \alpha_1 & -(1 - \alpha_1) \\ 1 - \alpha_0 - \alpha_1 & 1 - \alpha_0 \end{pmatrix}$  is called the discount matrix and can be written in the general form:

$$D = F - gw'. \quad (25)$$

The model is said to be stable if all the eigenvalues of (25) lie inside the unit circle. The eigenvalues are given by the following formula:

$$\lambda = \frac{2 - 2\alpha_0 + \alpha_1 \pm \sqrt{8\alpha_1 + 4\alpha_0 - 4\alpha_0\alpha_1 - 4 - 3\alpha_1^2}}{2}. \quad (26)$$

CES will be stable when the following system of inequalities is satisfied (see appendix E):

$$\begin{cases} (\alpha_0 - 2.5)^2 + \alpha_1^2 > 1.25 \\ (\alpha_0 - 0.5)^2 + (\alpha_1 - 1)^2 > 0.25 \\ (\alpha_0 - 1.5)^2 + (\alpha_1 - 0.5)^2 < 1.5 \end{cases} . \quad (27)$$

Both the stationarity and stability regions are shown in Figure 3. The stationarity region (23) corresponds to the triangle. All the combinations of smoothing parameters lying below the curve in the triangle will produce the stationary harmonic trajectories, while the rest lead to the exponential trajectories. The stability condition (27) corresponds to the dark region, which has an unusual form. The stability region intersects the stationarity region, but in general stable CES can produce both stationary and non-stationary forecasts.

#### 4.4 Conditional mean and variance of CES

The conditional mean of CES for  $h$  steps ahead with known  $l_t$  and  $c_t$  can be calculated using the state-space (20):

$$E(z_{t+h}|x_t) = F^{h-1}x_t, \quad (28)$$

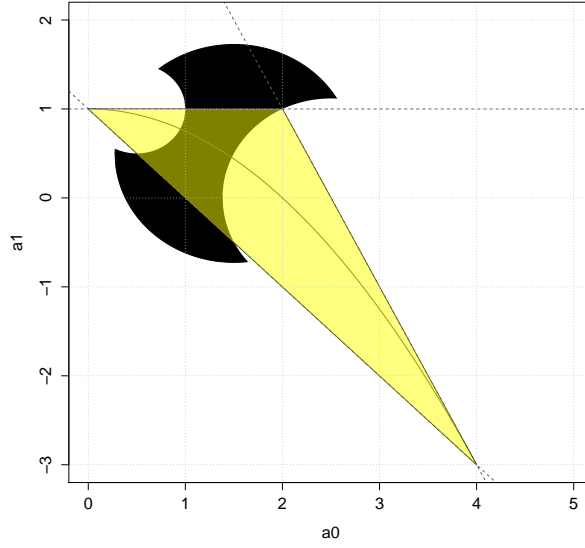
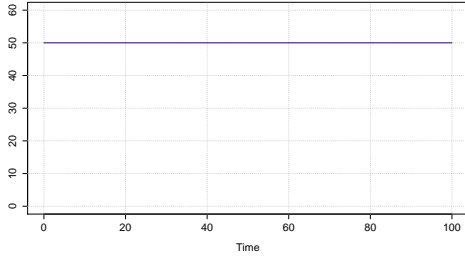


Figure 3: Stability and stationarity regions of CES, derived from the state-space form (20).

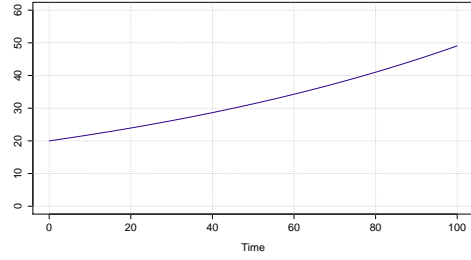
where  $E(z_{t+h}|x_t) = \hat{y}_{t+h} + i\hat{p}_{t+h}$  and  $F$  is the matrix from (8). The conditional mean estimated using (28) consists of two parts: the conditional mean of the actual values and the conditional mean of the information potential. The former is the main interest in forecasting.

The forecasting trajectories (28) will differ depending on the values of  $l_1$ ,  $c_1$  and the complex smoothing parameter. The analysis of stationarity conditions shows that there are several types of forecasting trajectories of CES depending on the particular value of the complex smoothing parameter:

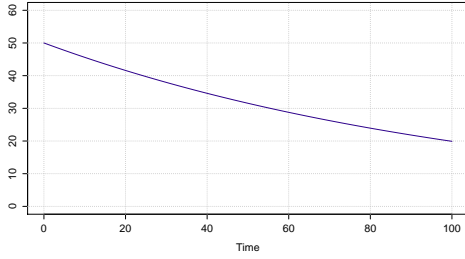
1. When  $\alpha_1 = 1$  all the values of the forecast will be equal to the last obtained forecast, which corresponds to a flat line. This trajectory is shown on the Figure 4a.
2. When  $\alpha_1 > 1$  the model produces the trajectory with the exponential growth which



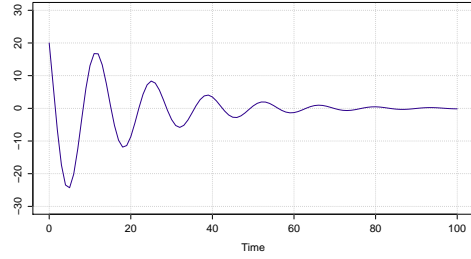
(a) the flat line



(b) exponential growth



(c) exponential decline



(d) harmonic oscillation

Figure 4: Forecasting trajectories.

is shown on Figure 4b.

3. When  $\frac{4-\alpha_0^2}{4} < \alpha_1 < 1$  the trajectory becomes stationary and CES produces the exponential decline shown on Figure 4c.
4. When  $1 - \alpha_0 < \alpha_1 < \frac{4-\alpha_0^2}{4}$  the trajectory becomes harmonic and will converge to zero (see Figure 4d).
5. Finally, when  $0 < \alpha_1 < 1 - \alpha_0$  the diverging harmonic trajectory is produced, the model becomes non-stationary. This forecasting trajectory is of no use in forecasting, that is why we do not show it on graphs.

Using (20) the conditional variance of CES for  $h$  steps ahead with known  $l_t$  and  $c_t$  can be calculated (Hyndman et al., 2008, p.96):

$$V(z_{t+h}|x_t) = \begin{cases} \sigma_\epsilon^2 \left( J_2 + \sum_{j=1}^h F^{j-1} g g' (F')^{j-1} \right) & \text{when } h > 1 \\ \sigma_\epsilon^2 & \text{when } h = 1 \end{cases}, \quad (29)$$

where  $J_2 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ .

The conditional variance estimated in (29) is a covariance matrix that contains the variance of the series, the variance of the information potential and the covariance between these two variables. However the variance of the actual series is of the main interest, so all the other elements can be ignored.

## 4.5 Connection with other forecasting models

### 4.5.1 Underlying ARMA model

Given the approximation (19) we can explore the connections with the other forecasting models further on. Model (14) transforms into ARMA(2,2) model:

$$\begin{cases} (1 - \phi_1 B - \phi_2 B^2)y_t = (1 - \theta_{1,1}B - \theta_{1,2}B^2)\epsilon_t \\ (1 - \phi_1 B - \phi_2 B^2)\xi_t = (1 - \theta_{2,1}B - \theta_{2,2}B^2)\epsilon_t \end{cases}, \quad (30)$$

where  $\phi_1 = 2 - \alpha_0$ ,  $\phi_2 = \alpha_0 + \alpha_1 - 2$ ,  $\theta_{1,1} = 2 - 2\alpha_0 + \alpha_1$ ,  $\theta_{1,2} = 3\alpha_0 + \alpha_1 - 2 - \alpha_0^2 - \alpha_1^2$ ,  $\theta_{2,1} = 2 + \alpha_1$  and  $\theta_{2,2} = \alpha_0 - \alpha_1 - 2$ .

The coefficients of AR terms of this model are connected with the coefficients of MA terms, via the complex smoothing parameters. This connection is non-linear but it imposes restrictions on the AR terms plane. Figure 5 demonstrates how the invertibility region restricts the AR coefficients field. The triangle on the plane corresponds to the stationarity

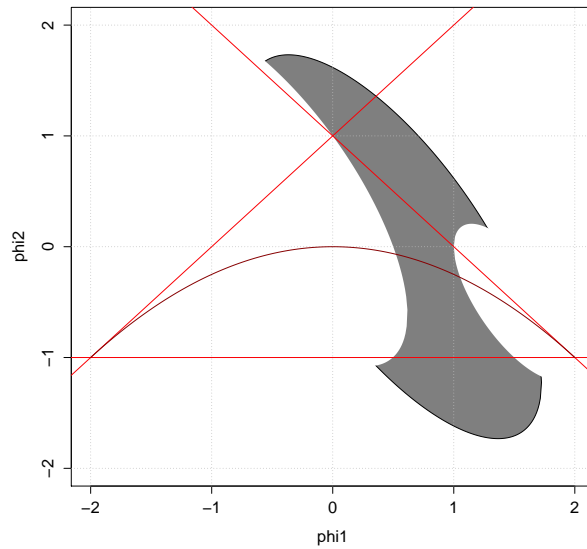


Figure 5: Invertibility region of CES on the plane of AR coefficients.

condition of AR(2) models, while the dark area demonstrates the invertibility region of CES.

Note that exponential smoothing models are nonstationary, but it is crucial for them to be stable or at least forecastable (Ord et al., 1997). In general the stability condition of ETS corresponds to the invertibility condition for ARIMA. That is why the later should be preferred to the former in CES framework. This means that CES will produce both stationary and non-stationary trajectories, depending on the complex smoothing parameter value. This selection between different types of processes happens naturally in the model without the need of model selection procedure. As an example, the similar stability condition on the same plane for ETS(A,N,N) corresponds to the dot with the coordinates (1,0) while for ETS(A,Ad,N) it corresponds to the line  $\phi_2 = 1 - \phi_1$  restricted by the segment  $\phi_1 \in (1, 2)$ .

### 4.5.2 Further observations on single exponential smoothing

Given (19) then several additional properties of CES become obvious after regrouping the elements of (6):

$$\begin{cases} \hat{y}_{t+1} = \hat{y}_t + \alpha_0 \epsilon_t - \alpha_1 p_t - (1 - \alpha_1) \hat{p}_t \\ \hat{p}_{t+1} = \hat{y}_t + \alpha_1 \epsilon_t + \alpha_0 p_t + (1 - \alpha_0) \hat{p}_t \end{cases}. \quad (31)$$

When  $\alpha_1$  is close to 1 the influence of  $\hat{p}_t$  on  $\hat{y}_{t+1}$  becomes minimal and the second smoothing parameter  $\alpha_0$  in (31) behaves similar to the smoothing parameter in SES:  $\alpha_0 - 1$  in CES becomes close to  $\alpha$  in SES. For example the value  $\alpha_0 = 1.24$  in CES will correspond to  $\alpha = 0.2$  in SES.

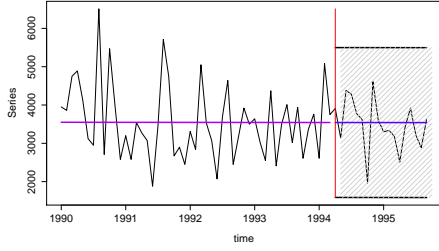
The insight from this is that when the series is stationary and the optimal smoothing parameter in SES should be close to zero, the optimal  $\alpha_0$  in CES will be close to one. At the same time the real part of the complex smoothing parameter will become close to 2 when the series demonstrates persistent changes in level.

## 5 Empirical results

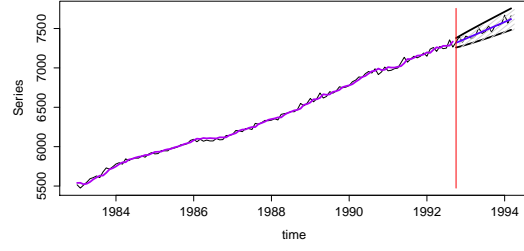
### 5.1 Real time series examples

We use CES on two time series to demonstrate how it performs on real data: a level series and a trend series from M3-Competition. The first series (number 1664) is shown in Figure 6a and the second series (number 2721) is shown in Figure 6b. Historic and fitted values, point and 95% interval forecasts produced by CES are plotted. Visual and statistical (using the ADF and KPSS tests) analysis indicate that series N1664 is stationary, while N2721 exhibits a clear trend.

Estimation of CES on the first time series results in the complex smoothing parameter



(a) Series number 1664



(b) Series number 2721

Figure 6: Stationary and trended time series. Red line divides the in-sample with the holdout-sample.

$\alpha_0 + i\alpha_1 = 0.99999 + 0.99996i$ . All the roots of the characteristic equation for such a complex smoothing parameter lie outside the unit circle and inequality (23) is satisfied which means that the model produces a stationary trajectory.

CES estimated on the second time series has complex smoothing parameter  $\alpha_0 + i\alpha_1 = 1.48187 + 1.00352i$ . This means that the forecast of CES on the second time series is influenced by a larger number of observations compared to the first time series. There are several roots of characteristic equation lying inside the unit circle and the imaginary part of the complex smoothing parameter is greater than one. All of this indicates that the model is non-stationary.

These two examples show that CES is capable of identifying whether the series is stationary or not and producing the appropriate forecast without the need for a model selection procedure.

## 5.2 M3 competition results

To evaluate the forecasting performance of CES against other exponential smoothing based methods we use the M3-competition dataset (Makridakis and Hibon, 2000) to conduct an empirical comparison. All 814 non-seasonal monthly time series are used. Table 1 lists the benchmarks used. These benchmarks have been selected to evaluate the performance of CES against various forms of exponential smoothing forecasts that are capable of capturing level and trend components. ETS(Z,Z,N) permits selecting the most appropriate non-seasonal model for each series, allowing any type of error or trend component.

The estimation of all the models was done in R. All the models excluding CES were estimated using “naive” and “ets” functions from “forecast” package by (Hyndman and Khandakar, 2008). A function “ces” from a package “CES” (<https://github.com/config-i1/CES>) was written by the authors of this paper to estimate CES.

We retain a similar setup to the original competition to facilitate comparability of the results. The last 18 observations are withheld as a test set that is forecasted by each of the methods considered. The forecasts are evaluated using the Mean Absolute Scaled Error (MASE) proposed by Hyndman and Koehler (2006).

Mean and median MASE over all the horizons are shown in the Table 2. The best performing method is highlighted in boldface. CES has a lowest mean and median MASE compared to all the other models used on this set of data. It outperforms ETS(A,N,N) and ETS(A,A,N) models, but also ETS(Z,Z,N) with the implemented model selection procedure.

To investigate why CES is more accurate than the other models mean and median MASE are calculated for each forecast horizon in Figure 7. The difference in accuracies between CES and other models increases with the increase of the forecasting horizon. For example, CES is substantially more accurate than the second best method ETS(M,Md,N)



Table 1: Forecasting methods used

Name	Method
Naive	Random walk model with Naive method
SES	ETS(A,N,N) which corresponds to Simple Exponential Smoothing method
AAN	ETS(A,A,N) which corresponds to Holt's method (Holt, 2004)
MMN	ETS(M,M,N) which underlies Pegel's method (Pegels, 1969)
AAAdN	ETS(A,Ad,N) which underlies additive damped trend method (Gardner and McKenzie, 1985)
MMdN	ETS(M,Md,N) which corresponds to multiplicative damped trend method (Taylor, 2003)
ZZN	ETS(Z,Z,N) – the general exponential smoothing model with the model selection procedure proposed by (Hyndman et al., 2002) using AICc.
CES	Complex Exponential Smoothing.

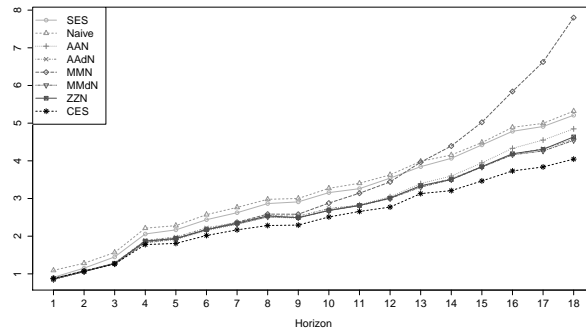
Table 2: MASE values for different forecasting models

Forecasting Method	Mean MASE	Median MASE
Naive	3.216	1.927
SES	3.098	1.681
AAN	2.755	1.602
MMN	3.323	1.703
AAAdN	2.726	1.467
MMdN	2.700	1.490
ZZN	2.716	1.466
<b>CES</b>	<b>2.496</b>	<b>1.387</b>

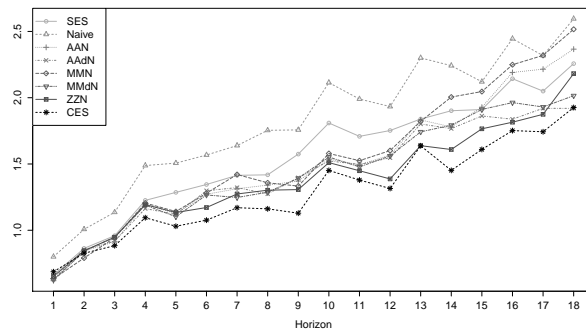
for horizons  $h = 6, \dots, 18$  but it is not the best method for the shorter horizons ( $h = 1, \dots, 3$ ). This indicates that CES was able to capture the long-term dependencies in time series better than the other exponential smoothing models.

## 6 Conclusions

In this paper we presented a new approach to time series modelling. Instead of taking into account only the actual value of series and decomposing it into several components, we introduced the notion of the information potential variable and combined it with the actual observations in one complex variable. Adopting this approach rather than using the arbitrary decomposition of time series into several components (level, trend, error) leads to a new model, the Complex Exponential Smoothing that is introduced in this paper.



(a) Mean MASE



(b) Median MASE

Figure 7: MASE values over different forecasting horizons for the competing models.

The state-space form using Single Source of Error was presented, which allowed the derivation of the stability and stationarity conditions of CES as well as conditional mean and variance of the model.

CES is a flexible model that can produce different types of trajectories. It encompasses both level and multiplicative trend series and approximates additive trend very well. One of the advantages of CES is that it smoothly moves from one trajectory to the other, depending on the complex smoothing parameter value. Furthermore, CES is able to distribute weights between different observations in time either exponentially or harmonically. This feature

allows CES to capture long-term dependencies and non-linear relations in time series.

CES has an underlying ARMAX(2,2) model, where the exogenous part corresponds to the information potential variable. By using the approximation discussed in the article, the model transforms into restricted ARMA(2,2). The main difference between ARMA(2,2) underlying CES and the general ARMA(2,2) is that in CES framework the AR and MA parameters become connected with each other. This leads to a different parameter space and a different modelling of time series.

Finally, CES is empirically shown to be more accurate than various exponential smoothing benchmarks, especially ETS(A,N,N), ETS(A,A,N) and ETS with model selection. This provides evidence that CES can be used effectively to capture both level and trend time series cases, side-stepping the model selection problem that conventional exponential smoothing and similar models face. We argue that the smooth transition of CES between level and trend series, in contrast to conventional ETS that implies abrupt changes due to the model selection between level and trend forms, is advantageous. CES is also able to capture long-term dependencies which results in more accurate forecasts for the longer horizons, in comparison to the exponential smoothing benchmarks considered here.

In conclusion, the Complex Exponential Smoothing model has unique and desirable properties for time series forecasting. Using the ideas of complex variables and information potential, CES builds on the established and widely used ideas behind exponential smoothing to overcome several limitations and modelling challenges of the latter.

## Acknowledgment

The authors would like to thank Keith Ord for his invaluable suggestions.

## A State-space form of CES

First of all any complex variable can be represented as a vector or as a matrix:

$$\begin{aligned} a + ib &= \begin{pmatrix} a \\ b \end{pmatrix}, \\ a + ib &= \begin{pmatrix} a & -b \\ b & a \end{pmatrix}. \end{aligned} \tag{32}$$

The general CES model (5) can be split into two parts: measurement and transition equations using (32):

$$\begin{cases} \begin{pmatrix} \hat{y}_t \\ \hat{x}_t \end{pmatrix} = \begin{pmatrix} l_{t-1} \\ c_{t-1} \end{pmatrix} \\ \begin{pmatrix} l_t \\ c_t \end{pmatrix} = \begin{pmatrix} \alpha_0 & -\alpha_1 \\ \alpha_1 & \alpha_0 \end{pmatrix} \begin{pmatrix} y_t \\ p_t \end{pmatrix} + \left( \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} - \begin{pmatrix} \alpha_0 & -\alpha_1 \\ \alpha_1 & \alpha_0 \end{pmatrix} \right) \begin{pmatrix} l_{t-1} \\ c_{t-1} \end{pmatrix} \end{cases}. \tag{33}$$

Regrouping the elements of transition equation in (33) the following equation can be obtained:

$$\begin{aligned} \begin{pmatrix} l_t \\ c_t \end{pmatrix} &= \begin{pmatrix} \alpha_0 & -\alpha_1 \\ \alpha_1 & \alpha_0 \end{pmatrix} \begin{pmatrix} 0 \\ p_t \end{pmatrix} + \begin{pmatrix} \alpha_0 & -\alpha_1 \\ \alpha_1 & \alpha_0 \end{pmatrix} \begin{pmatrix} y_t \\ 0 \end{pmatrix} + \\ &\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} l_{t-1} \\ c_{t-1} \end{pmatrix} - \begin{pmatrix} \alpha_0 & -\alpha_1 \\ \alpha_1 & \alpha_0 \end{pmatrix} \begin{pmatrix} 0 \\ c_{t-1} \end{pmatrix} - \begin{pmatrix} \alpha_0 & -\alpha_1 \\ \alpha_1 & \alpha_0 \end{pmatrix} \begin{pmatrix} l_{t-1} \\ 0 \end{pmatrix}. \end{aligned} \tag{34}$$

Grouping vectors of actual value and level component with complex smoothing parameter and then the level and information components leads to:

$$\begin{pmatrix} l_t \\ c_t \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} l_{t-1} \\ c_{t-1} \end{pmatrix} - \begin{pmatrix} 0 & -\alpha_1 \\ 0 & \alpha_0 \end{pmatrix} \begin{pmatrix} l_{t-1} \\ c_{t-1} \end{pmatrix} + \begin{pmatrix} \alpha_0 & -\alpha_1 \\ \alpha_1 & \alpha_0 \end{pmatrix} \begin{pmatrix} 0 \\ p_t \end{pmatrix} + \begin{pmatrix} \alpha_0 & -\alpha_1 \\ \alpha_1 & \alpha_0 \end{pmatrix} \begin{pmatrix} y_t - l_{t-1} \\ 0 \end{pmatrix}. \quad (35)$$

The difference between the actual value and the level in (35) is the error term:  $y_t - l_{t-1} = \epsilon_t$ . Using this and making several transformations gives the following state-space model:

$$\begin{cases} \begin{pmatrix} \hat{y}_t \\ \hat{x}_t \end{pmatrix} = \begin{pmatrix} l_{t-1} \\ c_{t-1} \end{pmatrix} \\ \begin{pmatrix} l_t \\ c_t \end{pmatrix} = \begin{pmatrix} 1 & -(1 - \alpha_1) \\ 1 & (1 - \alpha_0) \end{pmatrix} \begin{pmatrix} l_{t-1} \\ c_{t-1} \end{pmatrix} + \begin{pmatrix} -\alpha_1 \\ \alpha_0 \end{pmatrix} p_t + \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} \epsilon_t \end{cases}. \quad (36)$$

Now if CES should be represented in the state-space form with the SSOE then the measurement equation should also contain the same error term as the transition equation. Since the imaginary part of the measurement equation in (36) is unobservable, it does not contain any useful information for forecasting and can be excluded from the final state-space model:

$$\begin{cases} y_t = l_{t-1} + \epsilon_t \\ \begin{pmatrix} l_t \\ c_t \end{pmatrix} = \begin{pmatrix} 1 & -(1 - \alpha_1) \\ 1 & (1 - \alpha_0) \end{pmatrix} \begin{pmatrix} l_{t-1} \\ c_{t-1} \end{pmatrix} + \begin{pmatrix} -\alpha_1 \\ \alpha_0 \end{pmatrix} p_t + \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} \epsilon_t \end{cases}. \quad (37)$$

The other way of writing down this state-space model is by splitting the level and information components into two equations:

$$\begin{cases} y_t = l_{t-1} + \epsilon_t \\ l_t = l_{t-1} - (1 - \alpha_1)c_{t-1} - \alpha_1 p_t + \alpha_0 \epsilon_t \\ c_t = l_{t-1} + (1 - \alpha_0)c_{t-1} + \alpha_0 p_t + \alpha_1 \epsilon_t \end{cases} \quad (38)$$

## B Underlying ARIMA

The information component can be calculated using the second equation of (7) the following way:

$$c_{t-1} = -\frac{l_t - l_{t-1} + \alpha_1 p_t - \alpha_0 \epsilon_t}{1 - \alpha_1}. \quad (39)$$

Inserting (39) into the third equation of (7) leads to:

$$\frac{l_{t+1} - l_t + \alpha_1 p_{t+1} - \alpha_0 \epsilon_{t+1}}{1 - \alpha_1} = l_{t-1} - (1 - \alpha_0) \frac{l_t - l_{t-1} + \alpha_1 p_t - \alpha_0 \epsilon_t}{1 - \alpha_1} + \alpha_0 p_t + \alpha_1 \epsilon_t. \quad (40)$$

Multiplying both parts of (40) by  $-(1 - \alpha_1)$  and taking one lag back results in:

$$\begin{aligned} l_t - l_{t-1} + \alpha_1 p_t - \alpha_0 \epsilon_t = \\ -(1 - \alpha_1)l_{t-2} + (1 - \alpha_0)(l_{t-1} - l_{t-2} + \alpha_1 p_{t-1} - \alpha_0 \epsilon_{t-1}) - (1 - \alpha_1)\alpha_0 p_{t-1} - (1 - \alpha_1)\alpha_1 \epsilon_{t-1} \end{aligned} \quad (41)$$

Opening the brackets, transferring all the level components to the left hand side and all the error terms and information components to the right hand side and then regrouping the elements gives:

$$l_t - (2 - \alpha_0)l_{t-1} - (\alpha_0 + \alpha_1 - 2)l_{t-2} = \alpha_0 \epsilon_t - (\alpha_0 - \alpha_0^2 + \alpha_1 - \alpha_1^2)\epsilon_{t-1} - \alpha_1 p_t + (\alpha_1 - \alpha_0)p_{t-1}. \quad (42)$$

Now making substitutions  $l_t = y_{t+1} - \epsilon_{t+1}$  in (42), taking one more lag back and regrouping the error terms once again leads to:

$$\begin{aligned} y_t - (2 - \alpha_0)y_{t-1} - (\alpha_0 + \alpha_1 - 2)y_{t-2} = \\ \epsilon_t - (2 - 2\alpha_0)\epsilon_{t-1} - (2\alpha_0 + 2\alpha_1 - 2 - \alpha_0^2 - \alpha_1^2)\epsilon_{t-2} - \alpha_1 p_{t-1} - (\alpha_0 - \alpha_1)p_{t-2} \end{aligned} \quad (43)$$

The resulting model (43) is ARMAX(2,2) with lagged information potential:

$$(1 - \phi_1 B - \phi_2 B^2)y_t = (1 - \theta_1 B - \theta_2 B^2)\epsilon_t + (-\gamma_1 B - \gamma_2 B^2)p_t, \quad (44)$$

where  $\phi_1 = 2 - \alpha_0$ ,  $\phi_2 = \alpha_0 + \alpha_1 - 2$ ,  $\theta_1 = 2 - 2\alpha_0$ ,  $\theta_2 = 2\alpha_0 + 2\alpha_1 - 2 - \alpha_0^2 - \alpha_1^2$ ,  $\gamma_1 = \alpha_1$  and  $\gamma_2 = \alpha_0 - \alpha_1$ .

If the information potential is equal to the error term then the model (43) transforms into ARMA(2,2):

$$(1 - \phi_1 B - \phi_2 B^2)y_t = (1 - \theta_1 B - \theta_2 B^2)\epsilon_t, \quad (45)$$

where  $\phi_1 = 2 - \alpha_0$ ,  $\phi_2 = \alpha_0 + \alpha_1 - 2$ ,  $\theta_1 = 2 - 2\alpha_0 + \alpha_1$  and  $\theta_2 = 3\alpha_0 + \alpha_1 - 2 - \alpha_0^2 - \alpha_1^2$ .

In a similar manner using (7) it can be shown that the imaginary part of the series has the following underlying model:

$$\begin{aligned} c_t - (2 - \alpha_0)c_{t-1} - (\alpha_0 + \alpha_1 - 2)c_{t-2} = \\ \alpha_1 \epsilon_t - (\alpha_1 - \alpha_0)\epsilon_{t-1} + \alpha_0 p_t - (\alpha_0 + \alpha_1)p_{t-1} \end{aligned}, \quad (46)$$

If  $p_t - c_{t-1} = \xi_t$ , where  $\xi_t$  is an information gap, then the equation (46) can be rewritten:

$$\begin{aligned} p_{t+1} - (2 - \alpha_0)p_t - (\alpha_0 + \alpha_1 - 2)p_{t-1} - \alpha_0 p_t + (\alpha_0 + \alpha_1)p_{t-1} = \\ \xi_{t+1} - (2 - \alpha_0)\xi_t - (\alpha_0 + \alpha_1 - 2)\xi_{t-1} + \alpha_1 \epsilon_t - (\alpha_1 - \alpha_0)\epsilon_{t-1} \end{aligned}, \quad (47)$$



which after taking one lag back leads to the following simpler ARMAX(2,2) model:

$$(1 - \phi_1 B - \phi_2 B^2)\xi_t = (1 - \theta_1 B - \theta_2 B^2)p_t + (-\gamma_1 B - \gamma_2 B^2)\epsilon_t, \quad (48)$$

where  $\phi_1 = 2 - \alpha_0$ ,  $\phi_2 = \alpha_0 + \alpha_1 - 2$ ,  $\theta_1 = 2$ ,  $\theta_2 = 2$ ,  $\gamma_1 = \alpha_1$ , and  $\gamma_2 = \alpha_0 - \alpha_1$

In case of  $p_t = \epsilon_t$  the model transforms into the following ARMA(2,2):

$$(1 - \phi_1 B - \phi_2 B^2)\xi_t = (1 - \theta_1 B - \theta_2 B^2)\epsilon_t, \quad (49)$$

where  $\phi_1 = 2 - \alpha_0$ ,  $\phi_2 = \alpha_0 + \alpha_1 - 2$ ,  $\theta_1 = 2 + \alpha_1$  and  $\theta_2 = \alpha_0 - \alpha_1 - 2$ .

So CES has the following complex ARIMAX(2,0,2) model:

$$\begin{aligned} (1 - \phi_1 B - \phi_2 B^2)y_t &= (1 - \theta_{1,1} B - \theta_{1,2} B^2)\epsilon_t + (-\gamma_1 B - \gamma_2 B^2)p_t \\ (1 - \phi_1 B - \phi_2 B^2)\xi_t &= (1 - \theta_{2,1} B - \theta_{2,2} B^2)p_t + (-\gamma_1 B - \gamma_2 B^2)\epsilon_t \end{aligned}, \quad (50)$$

where  $\phi_1 = 2 - \alpha_0$ ,  $\phi_2 = \alpha_0 + \alpha_1 - 2$ ,  $\theta_{1,1} = 2 - 2\alpha_0$ ,  $\theta_{1,2} = 2\alpha_0 + 2\alpha_1 - 2 - \alpha_0^2 - \alpha_1^2$ ,  $\theta_{2,1} = 2$ ,  $\theta_{2,2} = 2$ ,  $\gamma_1 = \alpha_1$ ,  $\gamma_2 = \alpha_0 - \alpha_1$ .

## C The connection of CES and ETS(A,N,N)

When  $p_t = 0$  and  $\alpha_1 = 1$  the following state-space model based on (7) can be obtained:

$$\begin{cases} y_t = l_{t-1} + \epsilon_t \\ l_t = l_{t-1} + \alpha_0 \epsilon_t \\ c_t = l_{t-1} + (1 - \alpha_0)c_{t-1} + \epsilon_t \end{cases}. \quad (51)$$

The measurement equation in (51) implies that the information component  $c_t$  is constant over time, so the substitution  $c_{t-1} = c_t$  can be made and after opening the brackets and making several simple substitutions the following model is obtained:

$$\begin{cases} y_t = l_{t-1} + \epsilon_t \\ l_t = l_{t-1} + \alpha_0 \epsilon_t \\ c_t = \frac{l_{t-1}}{\alpha_0} + \frac{\epsilon_t}{\alpha_0} \end{cases} \quad (52)$$

## D Stationarity condition for CES

The analysis of the equation (22) shows that the eigenvalues can be either real or complex. In the cases of the real eigenvalues they need to be less than one, so the corresponding forecasting trajectory can be stationary and exponentially decreasing. This means that the following condition must be satisfied:

$$\begin{cases} \left| \frac{2 - \alpha_0 \pm \sqrt{\alpha_0^2 + 4\alpha_1 - 4}}{2} \right| < 1 \\ \alpha_0^2 + 4\alpha_1 - 4 > 0 \end{cases} \quad (53)$$

The first inequality in (53) leads to the following system of inequalities:

$$\begin{cases} \sqrt{\alpha_0^2 + 4\alpha_1 - 4} > \alpha_0 - 4 \\ \sqrt{\alpha_0^2 + 4\alpha_1 - 4} < \alpha_0 \\ -\sqrt{\alpha_0^2 + 4\alpha_1 - 4} > \alpha_0 - 4 \\ -\sqrt{\alpha_0^2 + 4\alpha_1 - 4} < \alpha_0 \end{cases} \quad (54)$$

The analysis of (54) shows that if  $\alpha_0 > 4$ , then the third inequality is violated and if  $\alpha_0 < 0$ , then the second inequality is violated. This means that the condition  $\alpha_0 \in (0, 4)$  is crucial for the stationarity of CES. This also means that the first and the fourth inequalities in (54) are always satisfied. Furthermore the second inequality can be transformed into:

$$\alpha_0^2 + 4\alpha_1 - 4 < \alpha_0^2, \quad (55)$$

which after simple cancellations leads to:

$$\alpha_1 < 1. \quad (56)$$

The other important result follows from the third inequality in (54), which can be derived using the condition  $\alpha_0 \in (0, 4)$ :

$$\alpha_0^2 + 4\alpha_1 - 4 < (\alpha_0 - 4)^2, \quad (57)$$

which implies that:

$$\alpha_1 < 5 - 2\alpha_0, \quad (58)$$

Uniting all these condition and taking into account the second inequality in (53), CES will produce a stationary exponential trajectory when:

$$\begin{cases} 0 < \alpha_0 < 4 \\ \alpha_1 < 5 - 2\alpha_0 \\ \frac{4-\alpha_0^2}{4} < \alpha_1 < 1 \end{cases} \quad (59)$$

The other possible situation is When the first part of the inequality (53) is violated, which will lead to the complex eigenvalues, meaning that the harmonic forecasting trajectory is produced. CES still can be stationary if both eigenvalues in (22) have absolute values less than one, meaning that:

$$\sqrt{\Re(\lambda)^2 + \Im(\lambda)^2} < 1 \quad (60)$$

This means in its turn the satisfaction of the following condition:

$$\sqrt{\left(\frac{2-\alpha_0}{2}\right)^2 + \left(i\frac{\sqrt{|\alpha_0^2 + 4\alpha_1 - 4|}}{2}\right)^2} < 1 \quad (61)$$

or:

$$0 \leq \frac{4 + \alpha_0^2 - 4\alpha_0}{4} - \frac{\alpha_0^2 + 4\alpha_1 - 4}{2} < 1 \quad (62)$$

which simplifying leads to:

$$1 < \alpha_0 + \alpha_1 \leq 2$$

or:

$$\begin{cases} \alpha_1 > 1 - \alpha_0 \\ \alpha_1 \leq 2 - \alpha_0 \end{cases}$$

The full condition that leads to the harmonic stationary trajectory of CES is:

$$\begin{cases} \alpha_1 < \frac{4-\alpha_0^2}{4} \\ \alpha_1 > 1 - \alpha_0 \\ \alpha_1 \leq 2 - \alpha_0 \end{cases} \quad (63)$$

Now it can be noted that the first inequality in (63) can be rewritten as a difference of squares:

$$\alpha_1 < (2 - \alpha_0) \frac{(2 + \alpha_0)}{4} \quad (64)$$

Comparing the right hand part of (64) with the right hand side of the third inequality in (63) it can be shown that there is only one point, when both of these inequalities will lead to the same constraint: when  $\alpha_0 = 2$ . This is because the line  $\alpha_1 = 2 - \alpha_0$  is a tangent

line for the function  $\alpha_1 = (2 - \alpha_0)\frac{(2+\alpha_0)}{4}$  in this point. In all the other cases the right hand part of (64) will be less than the right hand side of the third inequality in (63), which in its turn means that (64) can be substituted by stricter condition:

$$\begin{cases} \alpha_1 < \frac{4-\alpha_0^2}{4} \\ \alpha_1 > 1 - \alpha_0 \end{cases} \quad (65)$$

Uniting (65) with (59) leads to the following general stationarity condition:

$$\begin{cases} 0 < \alpha_0 < 4 \\ \alpha_1 < 5 - 2\alpha_0 \\ \frac{4-\alpha_0^2}{4} < \alpha_1 < 1 \\ \alpha_1 < \frac{4-\alpha_0^2}{4} \\ \alpha_1 > 1 - \alpha_0 \end{cases} \quad (66)$$

The third and fourth conditions can be united while the second and fifth conditions are always satisfied when the first condition is met (because the corresponding lines of these inequalities have an intersection in the point  $\alpha_0 = 4$ ). So the following simpler condition can be used instead of (66):

$$\begin{cases} \alpha_1 < 5 - 2\alpha_0 \\ \alpha_1 < 1 \\ \alpha_1 > 1 - \alpha_0 \end{cases} \quad (67)$$

## E Stability condition for CES

The general ARMA(2,2) will be invertible when the following condition is satisfied:

$$\begin{cases} \theta_2 + \theta_1 < 1 \\ \theta_2 - \theta_1 < 1 \\ \theta_2 > -1 \\ \theta_2 < 1 \end{cases} \quad (68)$$

Following from subsection 4.5.1, inserting the parameters from the ARMA (30) underlying CES the following system of inequalities is obtained:

$$\begin{cases} 3\alpha_0 + \alpha_1 - 2 - \alpha_0^2 - \alpha_1^2 + 2 - 2\alpha_0 + \alpha_1 < 1 \\ 3\alpha_0 + \alpha_1 - 2 - \alpha_0^2 - \alpha_1^2 - 2 + 2\alpha_0 - \alpha_1 < 1 \\ 3\alpha_0 + \alpha_1 - 2 - \alpha_0^2 - \alpha_1^2 > -1 \\ 3\alpha_0 + \alpha_1 - 2 - \alpha_0^2 - \alpha_1^2 < 1 \end{cases} \quad (69)$$

After the cancellations and regrouping of elements the system (69) transforms into:

$$\begin{cases} -\alpha_0^2 + 5\alpha_0 - \alpha_1^2 - 5 < 0 \\ -\alpha_0^2 + \alpha_0 - \alpha_1^2 + 2\alpha_1 - 1 < 0 \\ -\alpha_0^2 + 3\alpha_0 - \alpha_1^2 + \alpha_1 - 1 > 0 \\ -\alpha_0^2 + 3\alpha_0 - \alpha_1^2 + \alpha_1 - 3 < 0 \end{cases} \quad (70)$$

The inequalities in (70) can be transformed into the inequalities, based on squares of differences:

$$\begin{cases} (\alpha_0 - 2.5)^2 + \alpha_1^2 > 1.25 \\ (\alpha_0 - 0.5)^2 + (\alpha_1 - 1)^2 > 0.25 \\ (\alpha_0 - 1.5)^2 + (\alpha_1 - 0.5)^2 < 1.5 \\ (\alpha_0 - 1.5)^2 + (\alpha_1 - 0.5)^2 > -0.5 \end{cases} \quad (71)$$

Note that any point on the plane of smoothing parameters satisfies the last inequality in (71), so it does not change the stability region and can be skipped.

## References

- Athanasopoulos, G. and A. de Silva (2012). Multivariate exponential smoothing for forecasting tourist arrivals. *Journal of Travel Research* 51(5), 640–652.
- Athanasopoulos, G., R. J. Hyndman, H. Song, and D. C. Wu (2011). The tourism forecasting competition. *International Journal of Forecasting* 27(3), 822 – 844. Special Section 1: Forecasting with Artificial Neural Networks and Computational Intelligence Special Section 2: Tourism Forecasting.
- Billah, B., M. L. King, R. D. Snyder, and A. B. Koehler (2006, April). Exponential smoothing model selection for forecasting. *International Journal of Forecasting* 22(2), 239–247.
- Brenner, J. L., D. A. D’Esopo, and A. G. Fowler (1968, November). Difference equations in forecasting formulas. *Management Science* 15(3), 141–159.
- Brown, R. (1956). Exponential smoothing for predicting demand. Cambridge, Massachusetts: Aurthur D. Little Inc.
- Fildes, R., M. Hibon, S. Makridakis, and N. Meade (1998, September). Generalising about univariate forecasting methods: Further empirical evidence. *International Journal of Forecasting* 14(3), 339–358.
- Gardner, E. S. (1985). Exponential smoothing: The state of the art. *Journal of Forecasting* 4(1), 1–28.

- Gardner, E. S. and J. Diaz-Saiz (2008). Exponential smoothing in the telecommunications data. *International Journal of Forecasting* 24(1), 170–174.
- Gardner, E. S. and E. McKenzie (1985). Forecasting trends in time series. *Management Science* 31(10), 1237–1246.
- Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting* 20(1), 5–10.
- Hyndman, R., M. Akram, and B. Archibald (2008). The admissible parameter space for exponential smoothing models. *Annals of the Institute of Statistical Mathematics* 60(2), 407–426.
- Hyndman, R. J. and Y. Khandakar (2008, 7). Automatic time series forecasting: The forecast package for r. *Journal of Statistical Software* 27(3), 1–22.
- Hyndman, R. J. and A. B. Koehler (2006, October). Another look at measures of forecast accuracy. *International Journal of Forecasting* 22(4), 679–688.
- Hyndman, R. J., A. B. Koehler, J. K. Ord, and R. D. Snyder (2008). *Forecasting With Exponential Smoothing: The State Space Approach*. Springer-Verlag Berlin Heidelberg.
- Hyndman, R. J., A. B. Koehler, R. D. Snyder, and S. Grose (2002, July). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting* 18(3), 439–454.
- Jose, V. R. R. and R. L. Winkler (2008, January). Simple robust averages of forecasts: Some empirical results. *International Journal of Forecasting* 24(1), 163–169.



- Kolassa, S. (2011). Combining exponential smoothing forecasts using akaike weights. *International Journal of Forecasting* 27(2), 238 – 251.
- Kourentzes, N., F. Petropoulos, and J. R. Trapero (2014). Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting* 30(2), 291 – 302.
- Maia, A. L. S. and F. de A.T. de Carvalho (2011). Holts exponential smoothing and neural network models for forecasting interval-valued time series. *International Journal of Forecasting* 27(3), 740 – 759.
- Makridakis, S. and M. Hibon (2000). The M3-competition: Results, conclusions and implications. *International Journal of Forecasting* 16(4), 451–476.
- Ord, K., A. B. Koehler, and R. D. Snyder (1997). Estimation and prediction for a class of dynamic nonlinear statistical models. *Journal of the American Statistical Association* 92(440), 1621–1629.
- Pegels, C. C. (1969). Exponential forecasting: Some new variations. *Management Science* 15(5), pp. 311–315.
- Svetunkov, S. (2012). *Complex-Valued Modeling in Economics and Finance*. SpringerLink : Bücher. Springer.
- Taylor, J. W. (2003). Exponential smoothing with a damped multiplicative trend. *International Journal of Forecasting* 19(4), 715–725.
- Wang, J.-J., J.-Z. Wang, Z.-G. Zhang, and S.-P. Guo (2012). Stock index forecasting based on a hybrid model. *Omega* 40(6), 758 – 766. Special Issue on Forecasting in Management Science.