Complex Exponential Smoothing

Ivan Svetunkov[a,*], Nikolaos Kourentzes[b], J. Keith Ord[c]

[a] *Centre for Marketing Analytics and Forecasting*
Lancaster University Management School, UK
[b] *Skövde Artificial Intelligence Lab*
University of Skövde, Sweden
[c] *McDonough School of Business*
Georgetown University, US

Abstract

Exponential smoothing has been one of the most popular forecasting methods used to support various decisions in organisations, in activities such as inventory management, scheduling, revenue management and other areas. Although its relative simplicity and transparency have made it very attractive for research and practice, identifying the underlying trend remains challenging with significant impact on the resulting accuracy. This has resulted in the development of various modifications of trend models, introducing a model selection problem. With the aim of addressing this problem, we propose the Complex Exponential Smoothing (CES), based on the theory of functions of complex variables. The basic CES approach involves only two parameters and does not require a model selection procedure. Despite these simplifications, CES proves to be competitive with, or even superior to existing methods. We show that CES has several advantages over conventional exponential smoothing models: it can model and forecast both stationary and non-stationary processes, and CES can capture both level and trend cases, as defined in the conventional exponential smoothing classification. CES is evaluated on several forecasting competition datasets, demonstrating better performance than established benchmarks. We conclude that CES has desirable features for time series modelling and opens new promising avenues for research.

[*] Correspondance: I. Svetunkov, Centre for Marketing Analytics and Forecasting, Lancaster University Management School, Lancaster, Lancashire, LA1 4YX, UK.

Email address: `i.svetunkov@lancaster.ac.uk` (Ivan Svetunkov)

Keywords: Forecasting, State Space Models, Exponential Smoothing, Complex variables

1. Introduction

Effective forecasting is an essential prerequisite to logistics and operations planning. Such forecasting exercises focus not only on expected levels of demand but also on the amount of stock required. This is true for various prescriptive analytics tasks. Ideally, a forecaster will produce a prediction distribution that may be used in planning decisions using, for example, a newsvendor model.

There are several important elements that must be considered in this process:

- There may be a large number of items to be considered;
- The available time series for any given item may be short;
- The planning process will often focus on the upper tail of the prediction distribution.

These factors influence the approach taken to forecasting in the following ways. A large number of time series, that can in practice reach even several hundred thousand (for example, in retailing Fildes et al., 2019), makes the automatic generation of forecast desirable. Forecasting systems typically offer some automatic model specification capabilities, where the performance and reliability of the specification and selection of forecasting models are crucial. The pool of models, from which this selection is done, depends on the length of time series, and the capabilities of the supporting software, among other factors. The limited data can restrict the pool to simpler models that have fewer parameters to estimate. Similarly, simple models are commonplace in practice as many companies still rely on spreadsheet software for any analysis that do not offer advanced statistical capabilities. These have contributed to the wide use of exponential smoothing in many application areas.

Even though the recent M5 forecasting competition (Makridakis et al., 2021) showed that Machine Learning (ML) can be very effective for short-term forecasting, these methods remain incompatible with many forecasting problems. ML methods require a substantial training sample, which in many cases translates to long time series. Moreover, even though advances in computing have increased the training speed of ML methods, identifying the correct specification and hyper-parameters remains a very computationally intensive and demanding process. This, together with the limited capacities of software in companies for advanced methods has stymied the spread of ML in practice. Further, the M5 forecasting competition results demonstrate that in terms of inventory performance, more complicated forecasting approaches do not differ from the simpler ones (Spiliotis et al., 2021). It should also be noted that a lot of the ML entries in the competition failed to out-perform even the most basic time series procedures, exemplifying the difficulty to correctly specify ML methods. Although we cannot use these findings to make a general statement, it does support the continued use of well-researched statistical methods in many business applications.

Nonetheless, when it comes to statistical models, they assume some error process that is independent of the signal. This is motivated by mathematical convenience but rarely holds in reality. Forecasting models will only approximate the process underlying real data, and although "All models are wrong, but some are useful" (G. Box), the information in forecast errors is often ignored. This suggests that the residuals generated by a model may contain useful information that can provide feedback to improve subsequent forecasts.

Given these observations, the purpose of this paper is to develop a forecasting procedure that is parsimonious in terms of the number of parameters to be estimated, does not require model selection, eliminating both the complexity and the potential errors it introduces, and enables any information in the error estimates to feed back into subsequent forecasts.

The rest of the article is structured as follows. We provide a literature review in Section 2, then we introduce the Complex Exponential Smoothing (CES) in Section 3. Within that section the properties of CES and connections with existing models are discussed. Finally, we benchmark CES against established statistical models using real cases in Section 4, which is then followed by concluding remarks.

2. Literature Review

Exponential smoothing is a very successful group of forecasting methods which has been widely explored in theoretical research (for examples see Brown et al., 1961; McKenzie, 1986; Chen et al., 2000; Kim and Ryan, 2003; Zhu and Thonemann, 2004; Jose and Winkler, 2008; Kolassa, 2011; Wang et al., 2012; Athanasopoulos and de Silva, 2012; Rostami-Tabar et al., 2013) and used in practice (Fildes et al., 1998; Makridakis and Hibon, 2000; Gardner and Diaz-Saiz, 2008; Athanasopoulos et al., 2011).

The exponential smoothing methods were well known and popular amongst practising forecasters for almost half a century, originating in the work by Brown (1956). Hyndman et al. (2002), based on work by Snyder (1985) and Ord et al. (1997), embedded exponential smoothing within a state space framework, providing its statistical rationale, resulting in the ETS model (short for "Error, Trend and Seasonality"), which supports 30 models with different types of Error, Trend and Seasonal components. These may be None (N), Additive (A), Multiplicative (M), and for the trend the letter d is used to signify a damped trend. For example, ETS(A,N,A) would denote the model with additive error term, no trend and additive seasonality. An interested reader is referred to the textbook of Hyndman et al. (2008). ETS provides a systematic framework to estimate parameters, construct prediction intervals and choose between different types of exponential smoothing. The model selection in ETS framework relies on information criteria, such as Akaike Information Criterion (Akaike, 1974):

$$AIC = 2k + 2\ell(\theta), \tag{1}$$

where k is the number of estimated parameters and $\ell(\theta)$ is the log-likelihood function calculated for the data based on the vector of parameters θ. The ETS approach implies that all the possible ETS models are fit to the data, and the one with the lowest AIC is selected and then used for forecasting. This allows for substantial improvements in automating its use, particularly in terms of parameter estimation and model form selection (Hyndman et al., 2008). The framework has been widely used since, in different modifications of exponential smoothing (Gould et al., 2008; De Livera et al., 2011; Koehler et al., 2012; Taylor and Snyder, 2012; Kourentzes et al., 2014; Guo et al., 2016).

While ETS is widely used, recent research has demonstrated that it is possible to improve upon it. The literature 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 forecast accuracy (Kolassa, 2011; Kourentzes et al., 2014), with apparent implications for practice. The results of Kolassa (2011) imply that selecting the most appropriate ETS model is a challenging task. We argue that on the one hand models in the existing taxonomy do not cover all the possible forms; and on the other hand, this large diversity of models complicates the selection of the appropriate one. A further 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, as the components are not observable and their interpretation is elusive. For example the level component is non-stationary and can produce long term increases in a time series, much like a trend, complicating the separation between the two. The ETS framework suggests the existence of 5 types of alternative trend components, and selecting the most appropriate can be a challenging task and highly depends on the selection methodology used. Although in the original introduction of exponential smoothing the separation between level and trend was advantageous (Holt, 2004), the subsequent introduction of diverse trend alternatives opened a more difficult discussion of model selection (Hyndman et al., 2002). Kourentzes et al. (2014) showed that the complexity in identifying the presence and type of trend often results in misidentification and poorly performing long-term forecasts. Barrow et al. (2020) provide evidence that even when the ETS components are appropriately identified, the estimated smoothing parameters and the implied weight distribution (see Section 3.1) can result in poor forecasts, suggesting these as areas for improvement. Our motivation in this paper is to try to avoid this artificial distinction.

We propose a new model that eliminates the arbitrary distinction between level and trend components, involves only two parameters and effectively sidesteps the model selection procedure. We encode the observed value of a time series together with the error term as a complex variable, 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 performance empirically, making the use of complex variables instrumental. The key advantage of CES is that it can model both stationary and non-stationary time series, while conventional ETS is restricted to non-stationary time series. Furthermore, CES smoothly transitions between the two cases, avoiding discontinuous changes, as imposed by the ARIMA modelling framework. We view the proposed CES as a promising extension to the established exponential smoothing framework, which given its prevalence in fields such as forecasting, in both research and practice, has direct implications for practice.

3. Complex Exponential Smoothing

3.1. Method and model

Using the complex valued representation of time series, we propose the Complex Exponential Smoothing (CES) in analogy to the conventional exponential smoothing methods. Consider the simple exponential smoothing method:

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

where α is the smoothing parameter and $\hat{y}_t$ is the estimated value of series. The same method can be represented as a weighted average of previous actual observations if we substitute $\hat{y}_{t-1}$ by the formula (2) with an index $t-1$ instead of t (Brown, 1956):

$$\hat{y}_t = \alpha \sum_{j=1}^{t-1} (1-\alpha)^{j-1} y_{t-j}. \tag{3}$$

The idea of this representation is to demonstrate how the weights $\alpha(1-\alpha)^{j-1}$ are distributed over time in our sample. If the smoothing parameter $\alpha \in (0,1)$ then the weights decline exponentially with the increase of j. If it lies in the so called "admissible bounds" (Brenner et al., 1968), that is $\alpha \in (0,2)$, then the weights decline in oscillating manner. Both traditional and admissible bounds have been used efficiently in practice and in academic literature (for application of the latter see for example Gardner and Diaz-Saiz, 2008; Snyder et al., 2017). However, in real life the distribution of weights can be more complex, with harmonic rather than exponential decline, meaning that some of the past observation might have more importance than the recent ones. In order to implement such distribution of weights, we build upon (3) and introduce complex dynamic interactions by substituting the real variables with the complex ones in (3). First, we substitute y_{t-j} by the complex variable $y_{t-j} + ie_{t-j}$, where e_t is the error term of the model and i is the imaginary unit (which satisfies the equation $i^2 = -1$). The idea behind this is to have the impact of both actual values and the error on each observation in the past on the final forecast. Second, we substitute α with a complex variable $\alpha_0 + i\alpha_1$ and 1 by $1+i$ to introduce the harmonically declining weights. Depending on the values of the complex smoothing parameter, the weights distribution will exhibit a variety of trajectories over time, including exponential, oscillating and harmonic. Finally, the result of multiplication of two complex numbers will be another complex number, so we substitute $\hat{y}_{t-j}$ with $\hat{y}_{t-j} + i\hat{e}_{t-j}$, where $\hat{e}_{t-j}$ is the proxy for the error term. The Complex Exponential Smoothing obtained as a result of this can be written as:

$$\hat{y}_t + i\hat{e}_t = (\alpha_0 + i\alpha_1) \sum_{j=1}^{t-1} (1 + i - (\alpha_0 + i\alpha_1))^{j-1} (y_{t-j} + ie_{t-j}). \tag{4}$$

Having arrived to the model with harmonically distributed weights, we can now move to the shorter form by substituting

$$\hat{y}_{t-1} + i\hat{e}_{t-1} = (\alpha_0 + i\alpha_1) \sum_{j=2}^{t-1} (1 + i - (\alpha_0 + i\alpha_1))^{j-1} (y_{t-j} + ie_{t-j})$$

in (4) to get:

$$\hat{y}_t + i\hat{e}_t = (\alpha_0 + i\alpha_1)(y_{t-1} + ie_{t-1}) + (1 - \alpha_0 + i - i\alpha_1)(\hat{y}_{t-1} + i\hat{e}_{t-1}). \tag{5}$$

Note that $\hat{e}_t$ is not interesting for the time series analysis and forecasting purposes, but is used as a vessel containing the information about the previous errors of the method. Having the complex variables instead of the real ones in (5), allows taking the exponentially weighted values of both actuals and the forecast errors. By changing the value of $\alpha_0 + i\alpha_1$, we can regulate what proportions of the actual and the forecast error should be carried out to the future in order to produce forecasts.

Representing the complex-valued function as a system of two real-valued functions leads to:

$$\begin{aligned} \hat{y}_t &= (\alpha_0 y_{t-1} + (1 - \alpha_0)\hat{y}_{t-1}) - (\alpha_1 e_{t-1} + (1 - \alpha_1)\hat{e}_{t-1}) \\ \hat{e}_t &= (\alpha_1 y_{t-1} + (1 - \alpha_1)\hat{y}_{t-1}) + (\alpha_0 e_{t-1} + (1 - \alpha_0)\hat{e}_{t-1}) \end{aligned}. \tag{6}$$

CES introduces an interaction between the real and imaginary parts, and the equations in (6) are connected via the previous values of each other, causing interactions over time, defined by complex smoothing parameter value.

But the method itself is restrictive and does not allow easily producing prediction intervals and deriving the likelihood function. It is also important to understand what sort of statistical model underlies CES. This model can be written in the following state space form (see Appendix A for the derivation):

$$\begin{aligned} y_t &= l_{t-1} + \epsilon_t \\ l_t &= l_{t-1} - (1 - \alpha_1) c_{t-1} + (\alpha_0 - \alpha_1)\epsilon_t , \\ c_t &= l_{t-1} + (1 - \alpha_0) c_{t-1} + (\alpha_0 + \alpha_1)\epsilon_t \end{aligned} \tag{7}$$

where ϵ_t is the white noise error term, l_t is the level component and c_t is the non-linear trend component at observation t. Observe that dependencies in time series have an interactive 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. Although we call the c_t component as "non-linear trend", it does not correspond to the conventional trend component, because it contains the information of both previous c_{t-1} and the level l_{t-1}. Also, note that we use ϵ_t instead of e_t in (7), which means that the CES has (7) as an underlying statistical model only when there is no misspecification error. In the case of the estimation of this model, the ϵ_t will be substituted by e_t, which will then lead us to the original formulation (5).

This idea allows rewriting (7) in a shorter more generic way, resembling the general Single Source of Error (SSOE) state space framework:

$$\begin{aligned} y_t &= \mathbf{w}'\mathbf{v}_{t-1} + \epsilon_t \\ \mathbf{v}_t &= \mathbf{F}\mathbf{v}_{t-1} + \mathbf{g}\epsilon_t \end{aligned} , \tag{8}$$

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

The state space form (8) permits extending 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 transition matrix in (8) includes smoothing parameters which is not a standard feature of ETS models. Furthermore persistence vector includes the interaction of complex smoothing parameters, rather than smoothing parameters themselves.

The error term in (7) is additive, so the likelihood function for CES is trivial and is similar to the one in the additive exponential smoothing models

(Hyndman et al., 2008, p.68):

$$\mathcal{L}(\mathbf{g}, \mathbf{v}_0, \sigma^2 | \mathbf{Y}) = \left(\frac{1}{\sigma \sqrt{2\pi}} \right)^T \exp \left(-\frac{1}{2} \sum_{t=1}^{T} \left(\frac{\epsilon_t}{\sigma} \right)^2 \right), \tag{9}$$

where $\mathbf{v}_0$ is the vector of initial states, σ^2 is the variance of the error term and $\mathbf{Y}$ is the vector of all the in-sample observations.

3.2. Stationarity and stability conditions for CES

In order to understand the properties of CES, we need to study its stationarity and stability conditions. The former holds for general exponential smoothing in the state space form (8) when all the eigenvalues of $\mathbf{F}$ lie inside the unit circle (Hyndman et al., 2008, p.38). CES can be either stationary or not, depending on the complex smoothing parameter value, in contrast to ETS models that are always non-stationary. Calculating eigenvalues of $\mathbf{F}$ for CES gives the following roots:

$$\lambda = \frac{2 - \alpha_0 \pm \sqrt{\alpha_0^2 + 4\alpha_1 - 4}}{2}. \tag{10}$$

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 C):

$$\begin{cases} \alpha_1 < 5 - 2\alpha_0 \\ \alpha_1 < 1 \\ \alpha_1 > 1 - \alpha_0 \end{cases}. \tag{11}$$

Note that we are not restricting CES with the conditions (11), we merely show, how the model will behave depending on the value of the complex smoothing parameter. This property of CES means that it is able to model either stationary or non-stationary processes, without the need to switch between them. The property of CES for each separate time series depends on the value of the smoothing parameters.

The other important property that arises from (7) is the stability condition for CES. With $\epsilon_t = y_t - l_{t-1}$ the following is obtained:

$$\begin{aligned} & y_t = l_{t-1} + \epsilon_t \\ & \begin{pmatrix} l_t \\ c_t \end{pmatrix} = \begin{pmatrix} 1-\alpha_0+\alpha_1 & -(1-\alpha_1) \\ 1-\alpha_0-\alpha_1 & 1-\alpha_0 \end{pmatrix} \begin{pmatrix} l_{t-1} \\ c_{t-1} \end{pmatrix} + \begin{pmatrix} \alpha_0-\alpha_1 \\ \alpha_1+\alpha_0 \end{pmatrix} y_t \end{aligned} . \tag{12}$$

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

$$\mathbf{D} = \mathbf{F} - \mathbf{g}\mathbf{w}'. \tag{13}$$

The model is said to be stable if all the eigenvalues of (13) lie inside the unit circle. This is more important condition than the stationarity for the model, because it ensures that the complex weights decline over time and that the older observations have smaller weights than the new ones, which is one of the main features of the conventional ETS models. The eigenvalues are given by the following formula:

$$\lambda = \frac{2 - 2\alpha_0 + \alpha_1 \pm \sqrt{8\alpha_1 + 4\alpha_0 - 4\alpha_0\alpha_1 - 4 - 3\alpha_1^2}}{2}. \tag{14}$$

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

$$\begin{cases} (\alpha_0 - 2.5)^2 + \alpha_1^2 > 1.25 \\ (\alpha_0 - 0.5)^2 + (\alpha_1 - 1)^2 > 0.25 \\ (\alpha_0 - 1.5)^2 + (\alpha_1 - 0.5)^2 < 1.5 \end{cases} . \tag{15}$$

Both the stationarity and stability regions are shown in Figure 1. The stationarity region (11) 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 (15) corresponds to the dark region. The stability region intersects the stationarity region, but in general stable CES can produce both stationary and non-stationary forecasts.

3.3. 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 model (7):

$$\mathrm{E}\left(y_{t+h} | \mathbf{v}_t\right) = \mathbf{w}'\mathbf{F}^{h-1}\mathbf{v}_t, \tag{16}$$

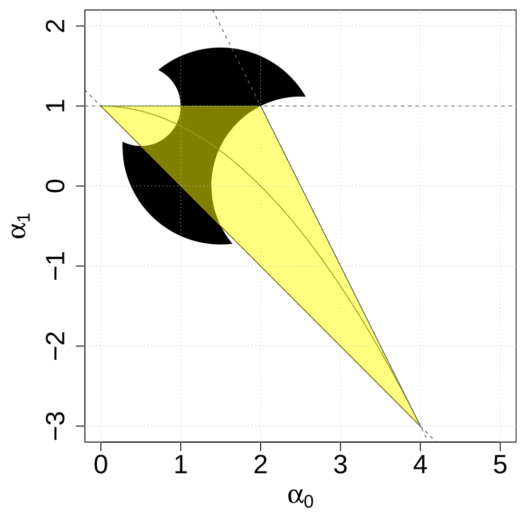


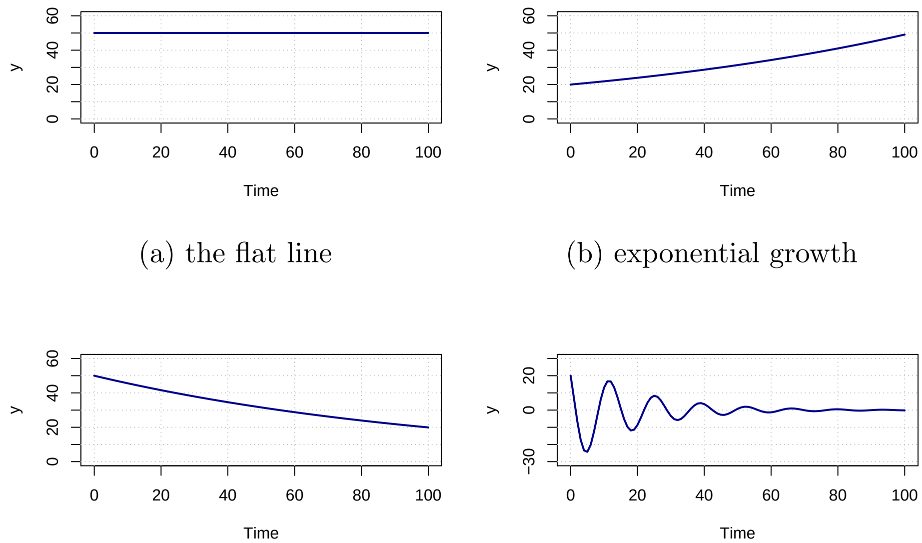
Figure 1: Stability (the black area) and stationarity (the light triangle) regions of CES, derived from the state space form (7).

where $\mathrm{E}(y_{t+h}|\mathbf{v}_t) = \hat{y}_{t+h}$, while $\mathbf{F}$ and $\mathbf{w}$ are the matrices from (8).

The forecasting trajectories of (16) will differ depending on the values of l_t, c_t and the complex smoothing parameter. The analysis of stationarity condition 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 forecast will be equal to the last obtained forecast, which corresponds to a flat line. This trajectory is shown in Figure 2a.
2. When $\alpha_1 > 1$ the model produces trajectory with exponential growth which is shown in Figure 2b.
3. When $\frac{4-\alpha_0^2}{4} < \alpha_1 < 1$ trajectory becomes stationary and CES produces exponential decline shown in Figure 2c.
4. When $1 - \alpha_0 < \alpha_1 < \frac{4-\alpha_0^2}{4}$ trajectory becomes harmonic and will converge to zero (see Figure 2d).
5. Finally, when $0 < \alpha_1 < 1 - \alpha_0$ the diverging harmonic trajectory is produced, the model becomes non-stationary. This trajectory is of no use in forecasting, that is why we do not show it on graphs.

Using (7) the conditional variance of CES for h steps ahead with known l_t and c_t can be calculated similarly to the pure additive ETS models (Hyndman et al., 2008, p.96).



(a) the flat line (b) exponential growth

(c) exponential decline (d) harmonic oscilation

Figure 2: Forecasting trajectories.

3.4. Connection with other forecasting models

3.4.1. Underlying ARMA model

All the pure additive 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 can be written in the form of ARMA(2,2) model (see Appendix B for the derivations):

$$\begin{cases} (1 - \phi_1 B - \phi_2 B^2) y_t = (1 - \theta_{1,1} B - \theta_{1,2} B^2)\epsilon_t \\ (1 - \phi_1 B - \phi_2 B^2) \xi_t = (1 - \theta_{2,1} B - \theta_{2,2} B^2)\epsilon_t \end{cases}, \tag{17}$$

where $\xi_t = \epsilon_t - c_{t-1}$, $\phi_1 = 2 - \alpha_0$, $\phi_2 = \alpha_0 + \alpha_1 - 2$, $\theta_{1,1} = 2 - 2\alpha_0 + \alpha_1$, $\theta_{1,2} = 3\alpha_0 + \alpha_1 - 2 - \alpha_0^2 - \alpha_1^2$, $\theta_{2,1} = -2 + \alpha_1$ 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 parameter. This connection is non-linear and imposes restrictions on the AR terms plane. Figure 3 demonstrates how the invertibility region restricts the AR coefficients field. The triangle on the plane corresponds to the stationarity condition of AR(2) models, while the dark area demonstrates the invertibility region of CES.

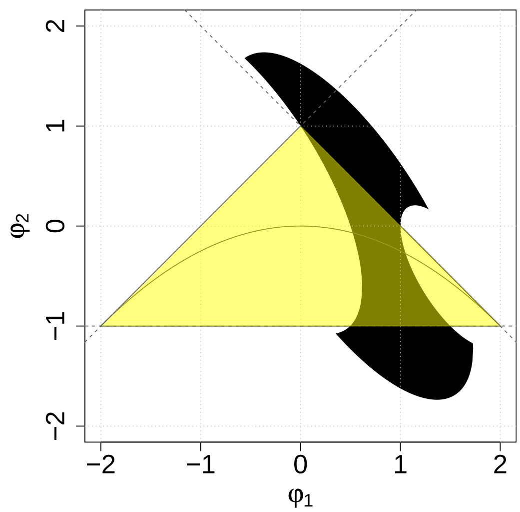


Figure 3: Invertibility (the black area) and stationarity (the light triangle) regions of CES on the plane of AR coefficients.

Note that exponential smoothing models are non-stationary in their nature, but it is crucial for them to be stable or at least forecastable (Ord et al., 1997). Thus we do not impose the stationarity condition on CES. Also, the stability condition of ETS corresponds to the invertibility condition for ARIMA, which is preferred in CES framework. This means that CES will produce both stationary and non-stationary trajectories, depending on the complex smoothing parameter value. This transition between different types of processes happens naturally in the model without the need of model selection procedure. For example, the similar stability condition on the same plane for ETS(A,N,N) corresponds to the point 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 $\phi_1 \in (1, 2)$.

3.4.2. Simple exponential smoothing (SES)

The additional properties of CES become obvious after regrouping the elements of (6):

$$\begin{cases} \hat{y}_t = \hat{y}_{t-1} + \alpha_0 e_{t-1} - \alpha_1 e_{t-1} - (1-\alpha_1)\hat{e}_{t-1} \\ \hat{e}_t = \hat{y}_{t-1} + \alpha_1 e_{t-1} + \alpha_0 e_{t-1} + (1-\alpha_0)\hat{e}_{t-1} \end{cases}. \tag{18}$$

When α_1 is close to 1 the influence of $\hat{e}_t$ on $\hat{y}_t$ becomes minimal and the second smoothing parameter α_0 in (18) behaves similarly to the smoothing parameter in SES: $\alpha_0 - 1$ in CES corresponds to α in SES. For example, the value $\alpha_0 = 1.24$ in CES will be equivalent to $\alpha = 0.24$ in SES.

Similar conclusions can be made by comparing ETS(A,N,N), which underlies SES, and CES in state space form (7), assuming that $\alpha_1 = 1$, which leads to:

$$\begin{aligned} y_t &= l_{t-1} + \epsilon_t \\ l_t &= l_{t-1} + (\alpha_0 - 1)\epsilon_t \\ c_t &= l_{t-1} + (1 - \alpha_0)c_{t-1} + (\alpha_0 + 1)\epsilon_t \end{aligned} . \tag{19}$$

The data generated using (19) will resemble the one generated using ETS(A,N,N). 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 α_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 corresponds to the random walk process.

3.5. Seasonal CES model

In order to make CES widely applicable, we also introduce a seasonal modification of the model. The simplest way to derive a seasonal model using CES is to use values of the level and non-linear trend components with a lag $t - m$ instead of $t - 1$. This is similar to the reduced seasonal exponential smoothing forms by Snyder and Shami (2001). In order to model both seasonal and non-seasonal parts we extend the original model (7) with a seasonal model, leading to the following seasonal CES model:

$$\begin{aligned} y_t &= l_{0,t-1} + l_{1,t-m} + \epsilon_t \\ l_{0,t} &= l_{0,t-1} - (1 - \alpha_1)c_{0,t-1} + (\alpha_0 - \alpha_1)\epsilon_t \\ c_{0,t} &= l_{0,t-1} + (1 - \alpha_0)c_{0,t-1} + (\alpha_0 + \alpha_1)\epsilon_t \ . \\ l_{1,t} &= l_{1,t-m} - (1 - \beta_1)c_{1,t-m} + (\beta_0 - \beta_1)\epsilon_t \\ c_{1,t} &= l_{1,t-m} + (1 - \beta_0)c_{1,t-m} + (\beta_0 + \beta_1)\epsilon_t \end{aligned} \tag{20}$$

Model (20) can still be written in a conventional state-space form (8). It exhibits several differences from conventional smoothing seasonal models. First, the proposed seasonal CES model in (20) does not have a set of usual seasonal components as the ordinary exponential smoothing models do, which means that there is no need to renormalise them. The values of $l_{1,t}$ and $c_{1,t}$ correspond to estimates of past level and non-linear trend components and have more common features with seasonal ARIMA (Box et al., 2015, p.300) than with the conventional seasonal exponential smoothing models. Second, it can be shown that the seasonal CES has an underlying model that corresponds to SARIMA$(2, 0, 2m+2)(2, 0, 0)_m$ (see Appendix E), which can be either stationary or not, depending on values of the complex smoothing parameters, similar to Subsection 3.4.1.

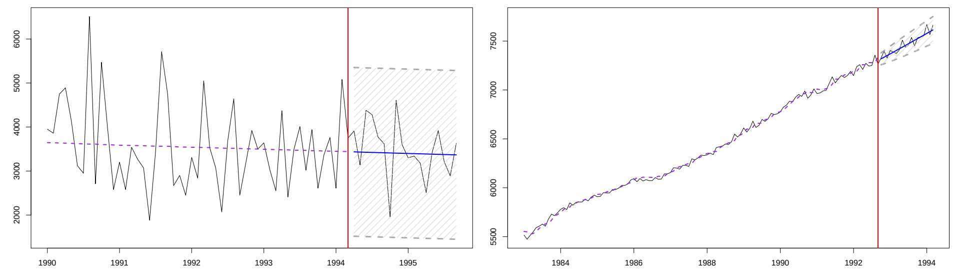
The seasonal CES can produce non-linear seasonality, i.e., the seasonal amplitude might increase or decreases in a non-linear fashion based on values of parameters, and all the possible types of trends discussed above, as the original level component $l_{0,t}$ can become negative while the lagged level component $l_{1,t}$ may become strictly positive. Furthermore, this model retains the property of independence of the original level and lagged level components, so it can model a multiplicative (or other) shape seasonality in the data even when the level of the series does not change. This could happen for example when the seasonality is either non-linear or when some other variable is determining its evolution, as for example in the case with solar power generation (Trapero et al., 2015).

When it comes to the estimation of this model, this can be done using the same principles as discussed earlier, while the stability condition can be checked by making sure that all the eigenvalues of the following discount matrix lie inside the unit circle:

$$\mathbf{D} = \begin{pmatrix} 1-\alpha_0+\alpha_1 & \alpha_1-1 & \alpha_1-\alpha_0 & 0 \\ 1-\alpha_0-\alpha_1 & 1-\alpha_0 & -\alpha_1-\alpha_0 & 0 \\ \beta_1-\beta_0 & 0 & 1-\beta_0+\beta_1 & \beta_1-1 \\ -\beta_1-\beta_0 & 0 & 1-\beta_0-\beta_1 & 1-\beta_0 \end{pmatrix} \tag{21}$$

The restrictions on parameters become more complicated than in the case of the non-seasonal model and cannot be easily analysed.

Finally, using the likelihood (9), the Akaike Information Criterion for both seasonal and non-seasonal CES models can be calculated based on formula (1). For the non-seasonal model (7), the number of estimated parameters k is equal to 5 (2 complex smoothing parameters, 2 initial states and 1 variance of the residuals), while, for the seasonal model, it becomes much greater than in the original model: $k = 4+2m+2+1$, which is 4 smoothing parameters, $2m$ initial lagged values, 2 initial values of the generic level and non-linear trend and 1 estimate of the variance. The flexibility of the model (20) comes at the cost of an increased number of parameters and AIC can help to identify whether this extra flexibility is beneficial or not. Naturally, other information criteria can be used.



(a) Series number 1664 (b) Series number 2721

Figure 4: Stationary and trended time series. The vertical line indicates the start of the holdout sample

Observe that the model selection problem for CES is reduced to choosing only between non-seasonal and seasonal variants, instead of the multiple model forms with conventional ETS.

4. Empirical results

In this section we provide two examples of the application of CES on real data, followed by a large empirical evaluation, comparing its performance against established time series model benchmarks.

4.1. Examples of application

To demonstrate the use of CES we use two real time series from the M3-Competition (Makridakis and Hibon, 2000). The first one (number 1664) is shown in Figure 4a and the second one (number 2721) is shown in Figure 4b. Graphical analysis, the Augmented Dickey–Fuller (ADF, Dickey and Fuller, 1979) and the Kwiatkowski– Phillips-Schmidt-Shin (KPSS, Kwiatkowski et al., 1992) tests suggest that the first series is stationary (Figure 4a), while the second one is non-stationary (Figure 4b), exhibiting a clear trend. The time series are split into in-sample and holdout subsets, as indicated in the figures with a vertical line.

Estimation of CES (using `ces()` function from the `smooth` package for R, Svetunkov, 2021b) on the first time series results in the complex smoothing parameter $\alpha_0 + i\alpha_1 = 0.99999 + 0.99884i$. All the roots of the characteristic equation for this complex smoothing parameter lie outside the unit circle and inequality (11) is satisfied, which means that the model produces a stationary trajectory. CES was able to identify that the series is stationary. The values of the parameters of ARMA, based on (17) are: $\phi_1 = 1.00001, \phi_2 = -0.00116, \theta_{1,1} = 0.99884, \theta_{1,2} = 0.00116$, which correspond to the stationary ARMA (the roots of the equation lie outside the unit circle) with long memory (the sum of the MA parameters is very close to one). Fitted values, point forecasts and 95% prediction intervals are provided in Figure 4a. All observations in the holdout sample lie in the 95% prediction interval.

For the second time series the complex smoothing parameter is $\alpha_0 + i\alpha_1 = 1.48098 + 1.00346i$. In this case, the forecast of CES 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, indicating that the model is non-stationary and that the bigger portion of the error term is taken into account when fitting the model. The respective values of the ARMA from (17) are: $\phi_1 = 0.51902, \phi_2 = 0.48444, \theta_{1,1} = 0.04150, \theta_{1,2} = 0.24616$. These values correspond to the non-stationary ARMA (the sum of AR parameters is greater than one) with a shorter memory (the sum of MA parameters is relatively small). Fitted values, point forecasts and the 95% prediction intervals are provided in Figure 4b.

These examples show that CES is capable of modelling both stationary and otherwise series, and that it can produce the appropriate forecasts without the need for a model selection procedure.

4.2. The experimental setup

In order to assess forecasting performance of CES we conduct an empirical evaluation on the M1 (Makridakis et al., 1982), M3 (Makridakis and Hibon, 2000) and Tourism (Athanasopoulos et al., 2011) competitions datasets. In total there are 5315 time series of yearly, quarterly, monthly and unidentified (‘other’ in the M3 dataset) frequency. The forecast horizons for the datasets correspond to those originally used in the competitions with 6 steps ahead for the yearly, 8 steps ahead for the quarterly and other, and 18 steps ahead for the monthly subsets.

We apply the `auto.ces()` function from the `smooth` package v3.1.2 (Svetunkov, 2021b, on CRAN) for R (R Core Team, 2021) [1]. This function implements the CES model with the proposed AIC based selection between

[1] Although we prefer to use R for data analysis, it should be noted that CES may also be implemented in Excel, with model fitting using Solver. As the only model selection

the seasonal and non-seasonal options. In order to see how CES performs, we also evaluate several benchmark methods:

1. ETS – the exponential smoothing model with the model selection procedure proposed by (Hyndman et al., 2002) using AICc implemented in `es()` function from `smooth` package for R with `model="ZXZ"`;
2. ARIMA – Automatic state space ARIMA (Svetunkov and Boylan, 2020) via `auto.ssarima()` from `smooth` package for R;
3. Theta – Theta method proposed by Assimakopoulos and Nikolopoulos (2000) and implemented in `thetaf()` function from `forecast` package v8.14 (Hyndman and Khandakar, 2008) for R;
4. SCUM - Simple Combination of Univariate Models proposed by Petropoulos and Svetunkov (2020), combining ETS, ARIMA, Theta and CES via medians;
5. SCUM(noCES) - As above but without including CES in the combination.

The first two are used to assess how CES compares with well established univariate benchmarks, while (3) is needed to see how CES compares with the winner of M3 competition. We also include the combination (4), which performed very well in the recent M4 competition (Makridakis et al., 2020) and (5) to see the role of CES in the combination.

The forecasting performance is evaluated using three error measures:

- RMSSE – Root Mean Squared Scaled Errors used in the M5 competition (Makridakis et al., 2022):

$$\text{RMSSE} = \bar{\Delta}_y^{-1} \sqrt{\frac{1}{h} \sum_{j=1}^{h} (y_{t+j} - \hat{y}_{t+j})^2}, \tag{22}$$

  where $\bar{\Delta}_y = \frac{1}{t-1} \sum_{j=2}^{t} |\Delta y_j|$ is the mean absolute value of the first differences $\Delta y_j = y_j - y_{j-1}$ of the in-sample data and h is the forecast horizon. We use this measure as it relies on the Mean Squared Error (MSE), which is minimised by the mean forecasts (Kolassa, 2020), which are produced by all the models under consideration.

question is between the seasonal and non-seasonal forms, implementation in spreadsheet is simpler than for ARIMA or ETS.

- MASE – Mean Absolute Scaled Error (Hyndman and Koehler, 2006):

$$\text{MASE} = \bar{\Delta}_y^{-1} \frac{1}{h} \sum_{j=1}^{h} |y_{t+j} - \hat{y}_{t+j}|, \tag{23}$$

  This is one of the standard error measures used in forecasting experiments. Relying on Mean Absolute Error (MAE), it is minimised by median rather than mean (Kolassa, 2020).

- MSIS – Mean Scaled Interval Score:

$$\begin{aligned} \text{MIS} =& \bar{\Delta}_y^{-1} \frac{1}{h} \sum_{j=1}^{h} \Bigg((u_{t+j} - l_{t+j}) + \frac{2}{\alpha}(l_{t+j} - y_{t+j}) \mathbb{1}\{y_{t+j} < l_{t+j}\} + \\ & \frac{2}{\alpha}(y_{t+j} - u_{t+j}) \mathbb{1}\{y_{t+j} > u_{t+j}\} \Bigg), \end{aligned} \tag{24}$$

  where $\mathbb{1}\{\cdot\}$ is the indicator function, u_{t+j} is the upper bound and l_{t+j} is the lower bound of the prediction interval and α is the significance level. We use 95% prediction interval in our estimation ($\alpha = 0.05$). This is a scale independent version of the Mean Interval Score (Gneiting and Raftery, 2007), and it shows the overall performance of the prediction intervals of models for selected quantiles.

We use the post-hoc Nemenyi test (Demšar, 2006), which relies on ranks of error measures. We use `rmcb()` function in the `greybox` package (Svetunkov, 2021a) for R. This reveals, whether there is evidence that the reported differences in accuracy between forecasts are statistically significant.

4.3. Results

Table 1 reports the summarised mean and median errors across all time series in the datasets. Contrasting the mean and the median values provides more information about the overall performance of models, also providing some insights on the distribution of error measures: for all RMSSE, MASE, and MSIS we observe large differences between the two, suggesting the presence of outlying errors. The best performing method in each column is highlighted in boldface. The combination approaches are distinguished from the other contestants and are compared separately.

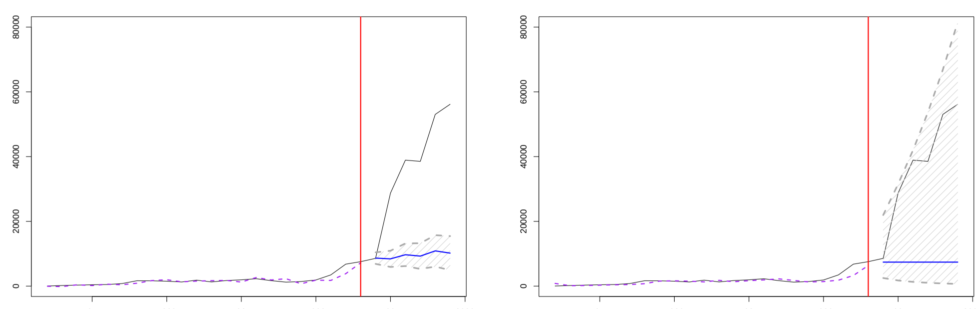
We see in Table 1 that CES outperforms other univariate models in terms of mean RMSSE and median values of RMSSE, MASE and MSIS. When it

Methods	Mean values			Median Values		
	RMSSE	MASE	MSIS	RMSSE	MASE	MSIS
CES	**1.959**	2.272	3.465	**1.170**	**1.298**	**0.869**
ETS	1.970	2.263	**2.258**	1.181	1.323	0.925
ARIMA	2.134	2.482	3.335	1.271	1.419	0.988
Theta	1.965	**2.252**	2.531	1.238	1.377	0.895
SCUM	**1.867**	**2.146**	2.333	**1.128**	**1.256**	**0.844**
SCUM(noCES)	1.911	2.191	**2.223**	1.143	1.279	0.859

Table 1: Error measures for M1, M3 and Tourism data

comes to the combination of models, SCUM does better in terms of mean and median RMSSE and MASE than SCUM(noCES) which does not included CES in the combination. This suggests that CES benefits the combination of the models, leading to improvements in accuracy. We argue that this happens because CES is able to capture long-term relations better than the other alternatives and can capture non-linear trends without a need to separate components into level and trend.

The comparison of mean and median MSIS in Table 1 demonstrates that CES does well in many cases, but fails in some, producing much higher errors than the other forecasting approaches. This is due primarily to one time series from M1 competition, where CES failed to identify the trend, and as a result the MSIS was equal to 523.034, while that for ETS was 25.590. This time series with CES and ETS is shown in Figure 5.



(a) CES applied to the data.

(b) ETS applied to the data.

Figure 5: Performance of CES ($\alpha_0 + i\alpha_1 = 2.108 + i1.118$) and ETS(M,N,N) ($\alpha = 0.881$) on Series 47 from M1 dataset.

As we can see from Figure 5, the series exhibits an unpredictable increase of values in the holdout. In our experiment, none of the models managed to capture it correctly, all of them underforecasted the data. However, the advantage of ETS in this situation was that the selected automatically ETS(M,N,N) model relied on multiplicative error and had a high smoothing parameter value, which resulted in an exploding prediction interval. While in many situations this would not be a suitable interval, but in this specific case it worked out better than the more conservative prediction interval of CES.

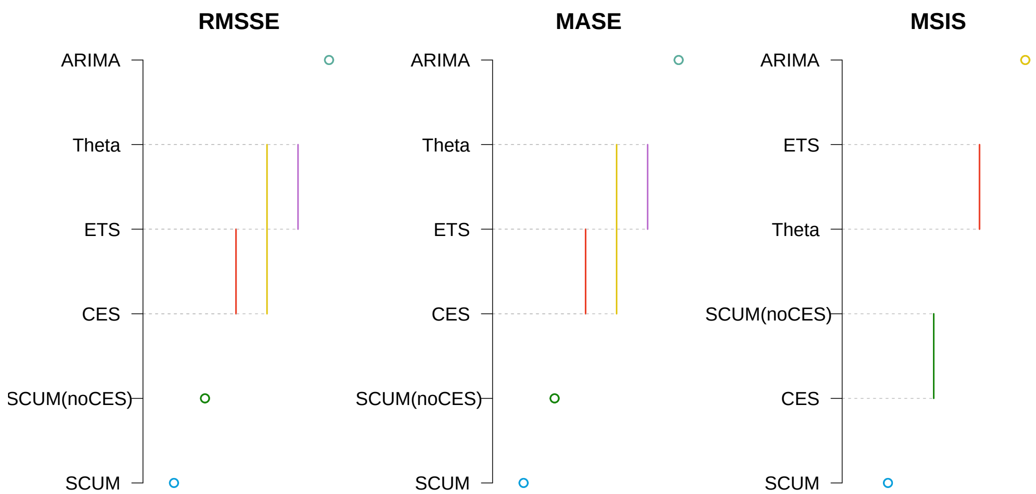


Figure 6: The Nemenyi test on RMSSE, MASE and MSIS for the datasets. The models are ordered based on their average ranks (the lower, the better).

In order to determine whether the observed accuracy differences between forecasts are statistically significant, we test them using the non-parametric Nemenyi tests. The results at 5% significance level are presented in Figure 6. In each panel of the Figure we rank the alternatives according to their mean rank (best at the bottom of the plot). For the forecasts that are connected with a vertical line there is no evidence of statistically significant differences at the 5% level. Note that there are multiple lines, depending on the forecast that one starts measuring from. Qualitatively the results for RMSSE and MASE are the same. CES ranks best when compared to all other univariate models, although the differences are not in all cases statistically significant. As expected, SCUM is significantly better than all the other approaches across all error measures, consistently outperforming SCUM(noCES). This means that CES does a significant contribution in the combination and is able to capture complex relations in the data that are not captured by the other forecasting models. Furthermore, when it comes to MSIS, CES ranks second, outperforming all the univariate models and the combination forecasts of SCUM(noCES).

Drawing overall conclusions from this experiment, CES is found to perform well across different error measures and its inclusion in combination forecasts leads to significant improvements across all error measures.

Having conducted the evaluation on M1, M3 and Tourism competition data, we also investigate the performance of the same set of models on the M4 forecasting competition data (Makridakis et al., 2020). The dataset contains a variety of time series and the models were assessed on different forecast horizons. Of particular interest is the large variety of sampling frequencies in the data. Details are presented in Table 2.

Type of data	Yearly	Quarterly	Monthly	Weekly	Daily	Hourly
Number of series	23,000	24,000	48,000	359	4,227	414
Forecast horizon	6	8	18	13	14	48

Table 2: Summary information of M4 data.

Several time series in the competition data exhibit extreme properties (see, Darin and Stellwagen, 2020; Ingel et al., 2020; Fildes, 2020), which caused the benchmark ETS to break down. This was due to influential outliers and structural breaks in the data. (Specifically, the time series 19357, 43160, 51552, 62738, 92370, 92435, 93159, 93196, 93412, 93491) These cases were excluded. The performance of models is summarised in Table 3.

Methods	Mean values			Median Values		
	MASE	RMSSE	MSIS	MASE	RMSSE	MSIS
CES	**2.921**	**2.339**	2.379	**1.803**	**1.529**	**0.592**
ETS	3.310	2.688	4.751	1.831	1.542	0.632
ARIMA	3.037	2.436	2.272	1.903	1.606	0.617
Theta	2.980	2.394	**2.130**	1.865	1.574	0.597
SCUM	**2.737**	**2.198**	**1.751**	**1.742**	**1.475**	**0.568**
SCUMnoCES	2.784	2.241	1.798	1.769	1.498	0.583

Table 3: Results from M4 competition, excluding series 19357, 43160, 51552, 62738, 92370, 92435, 93159, 93196, 93412, 93491.

As Table 3 demonstrates, CES outperforms all other models in terms of MASE and RMSSE. Furthermore, similar to the previous competition results, SCUM does consistently better than the combination without CES. All of this shows that CES brings value and performs well in practice.

4.4. Discussion

To further investigate the performance of CES, we analyse subsets of the previous experiment, focusing on specific data characteristics. First, we apply several special cases of ETS models to the non-seasonal data from the datasets above. To identify the non-seasonal series we rely on the model selection done by the ETS via the `es()` function from the `smooth` package for R. The results with 3262 time series are summarised in Table 4. Note that CES in this experiment is selected automatically between the seasonal and non-seasonal ones using AICc.

Methods	Mean values			Median Values		
	RMSSE	MASE	MSIS	RMSSE	MASE	MSIS
CES	*2.534*	*2.949*	4.269	**1.654**	*1.880*	*0.863*
ETS(ANN)	2.823	3.234	**3.104**	1.950	2.228	0.917
ETS(AAN)	2.670	3.095	3.785	1.694	1.919	0.907
ETS(AAdN)	**2.502**	**2.907**	3.769	*1.660*	**1.863**	**0.854**
ETS(MMN)	3.185	3.698	4.768	1.875	2.090	0.960
ETS(MMdN)	2.691	3.160	*3.546*	1.773	1.975	0.971

Table 4: Error measures for M1, M3 and Tourism data. Non-seasonal data. Comparison with special cases of ETS. The best results are in **boldface**, the second best are in *italic*.

We observe that CES substantially outperforms the local level, ETS(A,N,N), and local linear trend, ETS(A,A,N), models. The flexibility of CES allows it to capture different trajectories that are not explicitly captured by the conventional level and trend classification of exponential smoothing. Furthermore, it outperforms the multiplicative trend model, ETS(M,M,N), which produces exponential trajectories similar to the ones by CES, but is found to be less accurate than CES. This insight is in agreement with the two examples provided at the start of this section. It is lagging behind the damped trend model ETS(A,Ad,N) in terms of mean error measures, remaining second best in the majority of cases. The difference between the mean and median values for CES implies that it performed consistently well on the majority of series and failed on two of them, producing outlying errors.

In addition, we explore performance in the remaining seasonal time series (2053 time series). In this case, the ETS models allow for seasonality. Note that the split of the subset of the series was done according to ETS,

Methods	Mean values			Median Values		
	RMSSE	MASE	MSIS	RMSSE	MASE	MSIS
CES	**1.044**	**1.196**	2.188	**0.756**	**0.844**	**0.877**
ETS(ANA)	1.142	1.297	*2.037*	0.803	0.894	*0.920*
ETS(AAA)	1.135	1.295	2.247	0.802	0.877	0.934
ETS(AAdA)	*1.094*	*1.245*	2.155	0.797	0.876	0.946
ETS(MMM)	1.171	1.352	2.153	0.806	*0.875*	1.057
ETS(MMdM)	1.103	1.260	**1.934**	*0.789*	*0.875*	1.006

Table 5: Error measures for M1, M3 and Tourism data. Seasonal data. Comparison with special cases of ETS. The best results are in **boldface**, the second best are in *italic*.

making these the optimally selected models, given any trend restrictions we have imposed. The results of this experiment are shown in Table 5. CES outperforms all ETS models in terms of RMSSE and MASE. In terms of mean MSIS ETS(M,Md,M) ranks first, while CES outperforms the rest on median MSIS. This subset of results shows that the proposed use of information criteria can effectively identify when to switch between seasonal and non-seasonal CES.

5. Conclusions

In this paper we proposed a new approach to time series modelling. Instead of modelling only the observed value of series and decomposing it into several components, we use complex variables theory in order to connect the actual value and the forecast error. Our motivation in doing so is to provide the model with additional information about any potential misspecifications. Adopting this approach leads to the new model, the Complex Exponential Smoothing.

CES is a flexible model that can produce different types of forecast trajectories, avoiding the arbitrary decomposition of time series into several components (level, trend, error), which is the basis of the ETS family of models. It encompasses both level and multiplicative trend processes and approximates additive trend well. One of the advantages of CES is that it gradually transitions from one trajectory to the other, depending on its complex smoothing parameter value. In contrast to exponential smoothing, which inspired the creation of CES, it can model both stationary and non-stationary time series. This simplifies greatly the forecasting process, as model selection between these cases is reduced to simply estimating the complex smoothing parameter of CES. This also overcomes the arbitrary distinction between level and trend series done by conventional exponential smoothing and the resulting model selection.

We show that CES has the form of an ARMA(2,2) model, but without the stationarity or stability requirements. 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.

We also develop a seasonal counterpart of CES model to make it applicable to a wide variety of data. The seasonal CES is capable of modelling and forecasting time series with both additive, multiplicative, and other types of seasonality, providing even more flexibility to CES. We also discuss the selection between the seasonal and non-seasonal CES models, for which using information criteria is our recommendation.

Finally, CES is empirically shown to outperform ETS, ARIMA and Theta in terms of RMSSE. This validates that CES can be used efficiently to capture both level and trend time series, side-stepping the model selection problem in ETS, which forces abrupt changes between the various trend cases. The proposed model avoids this via a smooth transitions between them. We also evaluate the contribution of CES in the a combined forecast. We demonstrate that including CES in the forecast pool increases accuracy further. CES contributes to the combined forecasts due to its ability to capture long term trends and non-linear relations in the data.

We do not claim that CES is appropriate for every single forecasting scenario. Nonetheless, we provide evidence that it is a robust approach. In this work we have focused on the foundations of the basic idea, building upon the additive error model. Future research should investigate the impact of sample size on performance of CES and the extension of CES to incorporate a multiplicative error term. Furthermore, CES diverges from the traditional ETS structure, being able to capture a wider variety of trend and season behaviours and therefore could bring potential benefits. This is attributed to the usage of complex variables theory in the original method. Future research can investigate other options for models based on theory of complex variables.

Appendix A. State space form of CES

Any complex variable can be represented as a vector or as a matrix:

$$z = a + ib; \mathbf{z} = \begin{pmatrix} a \\ b \end{pmatrix}; \mathbf{z} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}. \tag{A.1}$$

The general CES model (5) can be split into two parts: measurement and transition equations using (A.1):

$$\begin{aligned} \begin{pmatrix} \hat{y}_t \\ \hat{e}_t \end{pmatrix} &= \begin{pmatrix} l_{t-1} \\ c_{t-1} \end{pmatrix} \\ \begin{pmatrix} l_t \\ c_t \end{pmatrix} &= \begin{pmatrix} \alpha_0 & -\alpha_1 \\ \alpha_1 & \alpha_0 \end{pmatrix} \begin{pmatrix} y_t \\ e_t \end{pmatrix} + \left(\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} - \begin{pmatrix} \alpha_0 & -\alpha_1 \\ \alpha_1 & \alpha_0 \end{pmatrix} \right) \begin{pmatrix} l_{t-1} \\ c_{t-1} \end{pmatrix} \end{aligned} \tag{A.2}$$

Regrouping the elements of transition equation in (A.2) the following equation can be obtained:

$$\begin{aligned} \begin{pmatrix} l_t \\ c_t \end{pmatrix} = & \begin{pmatrix} \alpha_0 & -\alpha_1 \\ \alpha_1 & \alpha_0 \end{pmatrix} \begin{pmatrix} 0 \\ e_t \end{pmatrix} + \begin{pmatrix} \alpha_0 & -\alpha_1 \\ \alpha_1 & \alpha_0 \end{pmatrix} \begin{pmatrix} y_t \\ 0 \end{pmatrix} \\ & + \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} l_{t-1} \\ c_{t-1} \end{pmatrix} - \begin{pmatrix} \alpha_0 & -\alpha_1 \\ \alpha_1 & \alpha_0 \end{pmatrix} \begin{pmatrix} 0 \\ c_{t-1} \end{pmatrix}. \\ & - \begin{pmatrix} \alpha_0 & -\alpha_1 \\ \alpha_1 & \alpha_0 \end{pmatrix} \begin{pmatrix} l_{t-1} \\ 0 \end{pmatrix} \end{aligned} \tag{A.3}$$

Grouping vectors of actual value and level component with complex smoothing parameter and then the level and non-linear trend components leads to:

$$\begin{aligned} \begin{pmatrix} l_t \\ c_t \end{pmatrix} = & \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} l_{t-1} \\ c_{t-1} \end{pmatrix} - \begin{pmatrix} 0 & -\alpha_1 \\ 0 & \alpha_0 \end{pmatrix} \begin{pmatrix} l_{t-1} \\ c_{t-1} \end{pmatrix} \\ & + \begin{pmatrix} \alpha_0 & -\alpha_1 \\ \alpha_1 & \alpha_0 \end{pmatrix} \begin{pmatrix} 0 \\ e_t \end{pmatrix} + \begin{pmatrix} \alpha_0 & -\alpha_1 \\ \alpha_1 & \alpha_0 \end{pmatrix} \begin{pmatrix} y_t - l_{t-1} \\ 0 \end{pmatrix} \end{aligned}. \tag{A.4}$$

The difference between the actual value and the level in (A.4) is the forecast error: $y_t - l_{t-1} = e_t$. If we consider the model that might generate the data for CES, then we should assume that there is no misspecification error, so that $e_t = \epsilon_t$. Using this and making several transformations gives the following state space model:

$$\begin{pmatrix} \hat{y}_t \\ \hat{\epsilon}_t \end{pmatrix} = \begin{pmatrix} l_{t-1} \\ c_{t-1} \end{pmatrix}$$
$$\begin{pmatrix} l_t \\ c_t \end{pmatrix} = \begin{pmatrix} 1 & -(1-\alpha_1) \\ 1 & (1-\alpha_0) \end{pmatrix} \begin{pmatrix} l_{t-1} \\ c_{t-1} \end{pmatrix} + \begin{pmatrix} -\alpha_1 \\ \alpha_0 \end{pmatrix} \epsilon_t + \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} \epsilon_t . \tag{A.5}$$

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 (A.5) is unobservable, it does not contain any useful information for forecasting and can be discarded from the final state space model:

$$\begin{aligned} y_t &= l_{t-1} + \epsilon_t \\ l_t &= l_{t-1} - (1-\alpha_1)c_{t-1} - \alpha_1\epsilon_t + \alpha_0\epsilon_t . \\ c_t &= l_{t-1} + (1-\alpha_0)c_{t-1} + \alpha_0\epsilon_t + \alpha_1\epsilon_t \end{aligned} \tag{A.6}$$

Appendix B. Underlying ARIMA

The non-linear trend component can be calculated using the second equation of (7), assuming $e_t = \epsilon_t$, in the following way:

$$c_{t-1} = -\frac{l_t - l_{t-1} + (\alpha_1 - \alpha_0)\epsilon_t}{1-\alpha_1}. \tag{B.1}$$

Inserting (B.1) into the third equation of (7) leads to:

$$-\frac{l_{t+1} - l_t + (\alpha_1 - \alpha_0)\epsilon_{t+1}}{1-\alpha_1} = l_{t-1} - (1-\alpha_0)\frac{l_t - l_{t-1} + (\alpha_1 - \alpha_0)\epsilon_t}{1-\alpha_1} + (\alpha_0+\alpha_1)\epsilon_t. \tag{B.2}$$

Multiplying both parts of (B.2) by $-(1-\alpha_1)$ and taking one lag back results in:

$$\begin{aligned} & l_t - l_{t-1} + (\alpha_1 - \alpha_0)\epsilon_t = \\ & -(1-\alpha_1)l_{t-2} + (1-\alpha_0)(l_{t-1} - l_{t-2} + (\alpha_1 - \alpha_0)\epsilon_{t-1}). \\ & -(1-\alpha_1)(\alpha_0+\alpha_1)\epsilon_{t-1} \end{aligned} \tag{B.3}$$

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

$$l_t - (2-\alpha_0)l_{t-1} - (\alpha_0+\alpha_1-2)l_{t-2} = (\alpha_0-\alpha_1)\epsilon_t - (\alpha_0-\alpha_0^2+\alpha_1-\alpha_1^2+\alpha_1-\alpha_0)\epsilon_{t-1}. \tag{B.4}$$

Now making substitutions $l_t = y_{t+1} - \epsilon_{t+1}$ in (B.4), taking one more lag back and regrouping the error terms once again leads to:

$$\begin{aligned} & y_t - (2-\alpha_0)y_{t-1} - (\alpha_0+\alpha_1-2)y_{t-2} = \\ & \epsilon_t - (2-2\alpha_0+\alpha_1)\epsilon_{t-1} - (3\alpha_0+\alpha_1-2-\alpha_0^2-\alpha_1^2)\epsilon_{t-2} \end{aligned}. \tag{B.5}$$

The resulting model (B.5) is ARMA(2,2):

$$(1-\phi_1 B - \phi_2 B^2)y_t = (1-\theta_{1,1}B - \theta_{1,2}B^2)\epsilon_t, \tag{B.6}$$

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

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

$$(1-\phi_1 B - \phi_2 B^2)\xi_t = (1-\theta_{2,1}B - \theta_{2,2}B^2)\epsilon_t, \tag{B.7}$$

where $\xi_t = \epsilon_t - c_{t-1}$, $\theta_{2,1} = 2+\alpha_1$ and $\theta_{2,2} = \alpha_0 - \alpha_1 - 2$.

Appendix C. Stationarity condition for CES

The analysis of the equation (10) 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:

$$\begin{cases} \left| \frac{2-\alpha_0 \pm \sqrt{\alpha_0^2+4\alpha_1-4}}{2} \right| < 1 \\ \alpha_0^2+4\alpha_1-4 \geq 0 \end{cases}. \tag{C.1}$$

The first inequality in (C.1) leads to the following system of inequalities:

$$\begin{cases} \sqrt{\alpha_0^2+4\alpha_1-4} > \alpha_0 - 4 \\ \sqrt{\alpha_0^2+4\alpha_1-4} < \alpha_0 \\ -\sqrt{\alpha_0^2+4\alpha_1-4} > \alpha_0 - 4 \\ -\sqrt{\alpha_0^2+4\alpha_1-4} < \alpha_0 \end{cases}. \tag{C.2}$$

The analysis of (C.2) 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 forth inequalities in (C.2) are always satisfied. Furthermore the second inequality can be transformed into:

$$\alpha_0^2 + 4\alpha_1 - 4 < \alpha_0^2, \tag{C.3}$$

which after simple cancellations leads to:

$$\alpha_1 < 1. \tag{C.4}$$

The other important result follows from the third inequality in (C.2), which can be derived using the condition $\alpha_0 \in (0, 4)$:

$$\alpha_0^2 + 4\alpha_1 - 4 < (\alpha_0 - 4)^2, \tag{C.5}$$

which implies that:

$$\alpha_1 < 5 - 2\alpha_0, \tag{C.6}$$

Uniting all these condition and taking into account the second inequality in (C.1), CES will produce a stationary exponential trajectory when:

$$\begin{cases} 0 < \alpha_0 < 4 \\ \alpha_1 < 5 - 2\alpha_0 \\ \dfrac{4 - \alpha_0^2}{4} \leq \alpha_1 < 1 \end{cases} . \tag{C.7}$$

The other possible situation is when the second part of the inequality (C.1) is violated, which will lead to the complex eigenvalues, meaning that the harmonic forecasting trajectory is produced. CES can still be stationary if both eigenvalues in (10) have absolute values less than one, meaning that:

$$\sqrt{\Re(\lambda)^2 + \Im(\lambda)^2} < 1, \tag{C.8}$$

where $\Re(\lambda)$ is the real part and $\Im(\lambda)$ is the imaginary part of λ. This means in its turn the satisfaction of the following condition:

$$\sqrt{\left(\frac{2 - \alpha_0}{2}\right)^2 + \left(i\frac{\sqrt{|\alpha_0^2 + 4\alpha_1 - 4|}}{2}\right)^2} < 1, \tag{C.9}$$

or:

$$0 \leq \frac{4+\alpha_0^2-4\alpha_0}{4} - \frac{\alpha_0^2+4\alpha_1-4}{4} < 1, \tag{C.10}$$

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:

$$\begin{cases} \alpha_1 < \frac{4-\alpha_0^2}{4} \\ \alpha_1 > 1-\alpha_0 \\ \alpha_1 \leq 2-\alpha_0 \end{cases}. \tag{C.11}$$

The first inequality in (C.11) can be rewritten as a difference of squares:

$$\alpha_1 < (2-\alpha_0)\frac{(2+\alpha_0)}{4}. \tag{C.12}$$

Comparing the right hand part of (C.12) with the right hand side of the third inequality in (C.11) 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 (C.12) will be less than the right hand side of the third inequality in (C.11). This means that the third inequality in (C.11) can be dropped:

$$\begin{cases} \alpha_1 < \frac{4-\alpha_0^2}{4} \\ \alpha_1 > 1-\alpha_0 \end{cases}. \tag{C.13}$$

Uniting (C.13) with (C.7) leads to the following general stationarity condition:

$$\begin{cases} 0 < \alpha_0 < 4 \\ \alpha_1 < 5 - 2\alpha_0 \\ \dfrac{4-\alpha_0^2}{4} \leq \alpha_1 < 1 \\ \alpha_1 < \dfrac{4-\alpha_0^2}{4} \\ \alpha_1 > 1 - \alpha_0 \end{cases}. \tag{C.14}$$

The third and fourth conditions can now be united. The first condition is always satisfied when conditions two and five are met (because the corresponding lines of these inequalities have an intersection in the point $\alpha_0 = 4$), so it can be removed. Finally, the following simpler condition can be used instead of (C.14):

$$\begin{cases} \alpha_1 < 5 - 2\alpha_0 \\ \alpha_1 < 1 \\ \alpha_1 > 1 - \alpha_0 \end{cases}. \tag{C.15}$$

Appendix D. Stability condition for CES

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

$$\begin{cases} \theta_2 + \theta_1 < 1 \\ \theta_2 - \theta_1 < 1 \\ \theta_2 > -1 \\ \theta_2 < 1 \end{cases}. \tag{D.1}$$

Due to (Hyndman et al., 2008, p.172 - 173) invertibility condition of ARIMA corresponds to stability condition of models in state space formulation. Following from subsection 3.4.1, inserting the parameters from ARMA (17) underlying CES, the following system of inequalities is obtained:

$$\begin{cases} 3\alpha_0 + \alpha_1 - 2 - \alpha_0^2 - \alpha_1^2 + 2 - 2\alpha_0 + \alpha_1 < 1 \\ 3\alpha_0 + \alpha_1 - 2 - \alpha_0^2 - \alpha_1^2 - 2 + 2\alpha_0 - \alpha_1 < 1 \\ 3\alpha_0 + \alpha_1 - 2 - \alpha_0^2 - \alpha_1^2 > -1 \\ 3\alpha_0 + \alpha_1 - 2 - \alpha_0^2 - \alpha_1^2 < 1 \end{cases}. \tag{D.2}$$

After the cancellations and regrouping of elements the system (D.2) transforms into:

$$\begin{cases} -\alpha_0^2 + 5\alpha_0 - \alpha_1^2 - 5 < 0 \\ -\alpha_0^2 + \alpha_0 - \alpha_1^2 + 2\alpha_1 - 1 < 0 \\ -\alpha_0^2 + 3\alpha_0 - \alpha_1^2 + \alpha_1 - 1 > 0 \\ -\alpha_0^2 + 3\alpha_0 - \alpha_1^2 + \alpha_1 - 3 < 0 \end{cases} . \tag{D.3}$$

The inequalities in (D.3) can be transformed into the inequalities, based on squares of differences:

$$\begin{cases} (\alpha_0 - 2.5)^2 + \alpha_1^2 > 1.25 \\ (\alpha_0 - 0.5)^2 + (\alpha_1 - 1)^2 > 0.25 \\ (\alpha_0 - 1.5)^2 + (\alpha_1 - 0.5)^2 < 1.5 \\ (\alpha_0 - 1.5)^2 + (\alpha_1 - 0.5)^2 > -0.5 \end{cases} . \tag{D.4}$$

Note that any point on the plane of smoothing parameters satisfies the last inequality in (D.4), so it is redundant and can be skipped.

Appendix E. General seasonal CES and SARIMA

The model (20) can be written in the following state-space form:

$$\begin{aligned} & y_t = \mathbf{w}_0' \mathbf{v}_{0,t-1} + \mathbf{w}_1' \mathbf{v}_{1,t-m} + \epsilon_t \\ & \mathbf{v}_{0,t} = \mathbf{F}_0 \mathbf{v}_{0,t-1} + \mathbf{g}_0 \epsilon_t \\ & \mathbf{v}_{1,t} = \mathbf{F}_1 \mathbf{v}_{1,t-m} + \mathbf{g}_1 \epsilon_t \end{aligned} , \tag{E.1}$$

where $\mathbf{v}_{0,t} = \begin{pmatrix} l_{0,t} \\ c_{0,t} \end{pmatrix}$ is the state vector of the non-seasonal part of CES, $\mathbf{v}_{1,t} = \begin{pmatrix} l_{1,t} \\ c_{1,t} \end{pmatrix}$ is the state vector of the seasonal part, $\mathbf{w}_0 = \mathbf{w}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ are the measurement vectors, $\mathbf{F}_0 = \begin{pmatrix} 1 & \alpha_1 - 1 \\ 1 & 1 - \alpha_0 \end{pmatrix}$, $\mathbf{F}_1 = \begin{pmatrix} 1 & \beta_1 - 1 \\ 1 & 1 - \beta_0 \end{pmatrix}$ are transition matrices and $\mathbf{g}_0 = \begin{pmatrix} \alpha_1 - \alpha_0 \\ \alpha_1 + \alpha_0 \end{pmatrix}$, $\mathbf{g}_1 = \begin{pmatrix} \beta_1 - \beta_0 \\ \beta_1 + \beta_0 \end{pmatrix}$ are persistence vectors of non-seasonal and seasonal parts respectively.

Observe that the lags of the non-seasonal and seasonal parts in (E.1) differ, which leads to splitting the state-space model into two parts. But uniting these parts will lead to the conventional state-space model:

$$\begin{aligned} y_t &= \mathbf{w}'\mathbf{v}_{t-\boldsymbol{l}} + \epsilon_t \\ \mathbf{v}_t &= \mathbf{F}\mathbf{v}_{t-\boldsymbol{l}} + \mathbf{g}\epsilon_t \end{aligned}, \tag{E.2}$$

where $\mathbf{v}_t = \begin{pmatrix} \mathbf{v}_{0,t} \\ \mathbf{v}_{1,t} \end{pmatrix}$, $\mathbf{v}_{t-\boldsymbol{l}} = \begin{pmatrix} \mathbf{v}_{0,t-1} \\ \mathbf{v}_{1,t-m} \end{pmatrix}$, $\mathbf{w} = \begin{pmatrix} \mathbf{w}_0 \\ \mathbf{w}_1 \end{pmatrix}$, $\mathbf{F} = \begin{pmatrix} \mathbf{F}_0 & 0 \\ 0 & \mathbf{F}_1 \end{pmatrix}$, $\mathbf{g} = \begin{pmatrix} \mathbf{g}_0 \\ \mathbf{g}_1 \end{pmatrix}$. The state vector $\mathbf{v}_{t-\boldsymbol{l}}$ can also be rewritten as $\mathbf{v}_{t-\boldsymbol{l}} = \begin{pmatrix} B & 0 \\ 0 & B^m \end{pmatrix} \begin{pmatrix} \mathbf{v}_{0,t} \\ \mathbf{v}_{1,t} \end{pmatrix}$, where B is a backshift operator. Making this substitution and taking $\mathbf{L} = \begin{pmatrix} B & 0 \\ 0 & B^m \end{pmatrix}$ the state-space model (E.2) can be transformed into:

$$\begin{aligned} y_t &= \mathbf{w}'\mathbf{L}\mathbf{v}_t + \epsilon_t \\ \mathbf{v}_t &= \mathbf{F}\mathbf{L}\mathbf{v}_t + \mathbf{g}\epsilon_t \end{aligned} \tag{E.3}$$

The transition equation in (E.3) can also be rewritten as:

$$(\mathbf{I}_2 - \mathbf{F}\mathbf{L})\mathbf{v}_t = \mathbf{g}\epsilon_t, \tag{E.4}$$

which after a simple manipulation leads to:

$$\mathbf{v}_t = (\mathbf{I}_2 - \mathbf{F}\mathbf{L})^{-1}\mathbf{g}\epsilon_t, \tag{E.5}$$

Substituting (E.5) into measurement equation in (E.3) gives:

$$y_t = \mathbf{w}'\mathbf{L}(\mathbf{I}_2 - \mathbf{F}\mathbf{L})^{-1}\mathbf{g}\epsilon_t + \epsilon_t. \tag{E.6}$$

Inserting the values of the vectors and multiplying the matrices leads to:

$$y_t = (1 + \mathbf{w}_0'(\mathbf{I}_2 - \mathbf{F}_0 B)^{-1}\mathbf{g}_0 B + \mathbf{w}_1'(\mathbf{I}_2 - \mathbf{F}_1 B^m)^{-1}\mathbf{g}_1 B^m)\epsilon_t. \tag{E.7}$$

Substituting the values by the matrices in (E.7) gives:

$$\begin{aligned} y_t = \Bigg(1 &+ \mathbf{w}_0' \begin{pmatrix} 1-B & (1-\alpha_1)B \\ -B & 1-B+\alpha_0 B \end{pmatrix}^{-1} \begin{pmatrix} \alpha_1 - \alpha_0 \\ \alpha_1 + \alpha_0 \end{pmatrix} B + \\ & \mathbf{w}_1' \begin{pmatrix} 1-B^m & (1-\beta_1)B^m \\ -B^m & 1-B^m+\beta_0 B^m \end{pmatrix}^{-1} \begin{pmatrix} \beta_1 - \beta_0 \\ \beta_1 + \beta_0 \end{pmatrix} B^m)\epsilon_t \Bigg). \end{aligned} \tag{E.8}$$

The inverse of the first matrix in (E.8) is equal to:

$$(\mathbf{I}_2 - \mathbf{F}_0 B)^{-1} = \frac{1}{1-2B-(\alpha_0+\alpha_1-2)B^2} \begin{pmatrix} 1-(1-\alpha_0)B & (\alpha_1-1)B \\ B & 1-B \end{pmatrix}, \tag{E.9}$$

similarly the inverse of the second matrix is:

$$(\mathbf{I}_2 - \mathbf{F}_1 B^m)^{-1} = \frac{1}{1-2B^m-(\beta_0+\beta_1-2)B^{2m}} \begin{pmatrix} 1-(1-\beta_0)B^m & (\beta_1-1)B^m \\ B^m & 1-B^m \end{pmatrix}. \tag{E.10}$$

Inserting (E.9) and (E.10) into (E.8), after cancellations and regrouping of elements leads to:

$$\begin{aligned} &(1-2B-(\alpha_0+\alpha_1-2)B^2)(1-2B^m-(\beta_0+\beta_1-2)B^{2m})y_t = \\ &\big[(1-2B-(\alpha_0+\alpha_1-2)B^2)(1-2B^m-(\beta_0+\beta_1-2)B^{2m})+ \\ &(1-2B^m-(\beta_0+\beta_1-2)B^{2m})(\alpha_1-\alpha_0-((\alpha_0-\alpha_1)^2-2\alpha_1)B)+ \\ &(1-2B-(\alpha_0+\alpha_1-2)B^2)(\beta_1-\beta_0-((\beta_0-\beta_1)^2-2\beta_1)B^m)\big]\,\epsilon_t \end{aligned} \tag{E.11}$$

Unfortunately, there is no way to simplify (E.11) to present it in a compact form, so the final model corresponds to SARIMA$(2,0,2m+2)(2,0,0)_m$.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19 (6), 716–723.

Assimakopoulos, V., Nikolopoulos, K., 2000. The theta model: a decomposition approach to forecasting. International Journal of Forecasting 16, 521–530.

Athanasopoulos, G., de Silva, A., 2012. Multivariate exponential smoothing for forecasting tourist arrivals. Journal of Travel Research 51 (5), 640–652.

Athanasopoulos, G., Hyndman, R. J., Song, H., Wu, D. C., 2011. The tourism forecasting competition. International Journal of Forecasting 27 (3), 822–844.

Barrow, D., Kourentzes, N., Sandberg, R., Niklewski, J., 2020. Automatic robust estimation for exponential smoothing: Perspectives from statistics and machine learning. Expert Systems with Applications 160, 113637.

Box, G. E., Jenkins, G. M., Reinsel, G. C., Ljung, G. M., 2015. Time series analysis: forecasting and control, 5th Edition. John Wiley & Sons.

Brenner, J. L., D'Esopo, D. A., Fowler, A. G., 1968. Difference equations in forecasting formulas. Management Science 15 (3), 141–159.

Brown, R. G., 1956. Exponential Smoothing for predicting demand. URL https://www.industrydocuments.ucsf.edu/docs/jzlc0130

Brown, R. G., Meyer, R. F., D'Esopo, D. A., 1961. The fundamental theorem of exponential smoothing. Operations Research 9 (5), 673–687.

Chen, F., Ryan, J. K., Simchi-Levi, D., 2000. The impact of exponential smoothing forecasts on the bullwhip effect. Naval Research Logistics (NRL) 47 (4), 269–286.

Darin, S. G., Stellwagen, E., 2020. Forecasting the M4 competition weekly data: Forecast Pro's winning approach. International Journal of Forecasting 36 (1), 135–141.

De Livera, A. M., Hyndman, R. J., Snyder, R. D., 2011. Forecasting time series with complex seasonal patterns using exponential smoothing. Journal of the American Statistical Association 106 (496), 1513–1527.

Demšar, J., 2006. Statistical comparisons of classifiers over multiple data sets. The Journal of Machine Learning Research 7, 1–30.

Dickey, D. A., Fuller, W. A., jun 1979. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. Journal of the American Statistical Association 74 (366a), 427–431.

Fildes, R., 2020. Learning from forecasting competitions. International Journal of Forecasting 36 (1), 186–188.

Fildes, R., Hibon, M., Makridakis, S., Meade, N., 1998. Generalising about univariate forecasting methods: Further empirical evidence. International Journal of Forecasting 14 (3), 339–358.

Fildes, R., Ma, S., Kolassa, S., 2019. Retail forecasting: Research and practice. International Journal of Forecasting (In Press). URL https://doi.org/10.1016/j.ijforecast.2019.06.004

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

Gardner, E. S., Diaz-Saiz, J., 2008. Exponential smoothing in the telecommunications data. International Journal of Forecasting 24 (1), 170–174.

Gneiting, T., Raftery, A. E., 2007. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102 (477), 359–378.

Gould, P. G., Koehler, A. B., Ord, J. K., Snyder, R. D., Hyndman, R. J., Vahid-Araghi, F., 2008. Forecasting time series with multiple seasonal patterns. European Journal of Operational Research 191 (1), 207–222.

Guo, X., Lichtendahl, K. C., Grushka-Cockayne, Y., 2016. Forecasting Life Cycles with Exponential Smoothing.
URL http://www.ssrn.com/abstract=2805244

Holt, C. C., 2004. Forecasting seasonals and trends by exponentially weighted moving averages. International Journal of Forecasting 20 (1), 5–10.

Hyndman, R. J., Khandakar, Y., 7 2008. Automatic time series forecasting: The forecast package for R. Journal of Statistical Software 27 (3), 1–22.

Hyndman, R. J., Koehler, A. B., 2006. Another look at measures of forecast accuracy. International Journal of Forecasting 22 (4), 679–688.

Hyndman, R. J., Koehler, A. B., Ord, J. K., Snyder, R. D., 2008. Forecasting With Exponential Smoothing: The State Space Approach. Springer-Verlag Berlin Heidelberg.

Hyndman, R. J., Koehler, A. B., Snyder, R. D., Grose, S., 2002. A state space framework for automatic forecasting using exponential smoothing methods. International Journal of Forecasting 18 (3), 439–454.

Ingel, A., Shahroudi, N., Kängsepp, M., Tättar, A., Komisarenko, V., Kull, M., 2020. Correlated daily time series and forecasting in the M4 competition. International Journal of Forecasting 36 (1), 121–128.

Jose, V. R. R., Winkler, R. L., Jan. 2008. Simple robust averages of forecasts: some empirical results. International Journal of Forecasting 24 (1), 163–169.

Kim, H.-K., Ryan, J. K., 2003. The cost impact of using simple forecasting techniques in a supply chain. Naval Research Logistics (NRL) 50 (5), 388–411.

Koehler, A. B., Snyder, R. D., Ord, J. K., Beaumont, A., 2012. A study of outliers in the exponential smoothing approach to forecasting. International Journal of Forecasting 28 (2), 477 – 484.

Kolassa, S., 2011. Combining exponential smoothing forecasts using Akaike weights. International Journal of Forecasting 27 (2), 238–251.

Kolassa, S., 2020. Why the best point forecast depends on the error or accuracy measure. International Journal of Forecasting 36 (1), 208–211.

Kourentzes, N., Petropoulos, F., Trapero, J. R., 2014. Improving forecasting by estimating time series structural components across multiple frequencies. International Journal of Forecasting 30 (2), 291–302.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., Shin, Y., 1992. Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root : How Sure Are We That Economic Time Series Are Nonstationary? Journal of Econometrics 54, 159–178.

Makridakis, S., Andersen, A. P., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., Winkler, R. L., 1982. The accuracy of extrapolation (time series) methods: Results of a forecasting competition. Journal of Forecasting 1 (2), 111–153.

Makridakis, S., Hibon, M., 2000. The M3-competition: Results, conclusions and implications. International Journal of Forecasting 16 (4), 451–476.

Makridakis, S., Spiliotis, E., Assimakopoulos, V., 2020. The M4 Competition: 100,000 time series and 61 forecasting methods. International Journal of Forecasting 36 (1), 54–74.

Makridakis, S., Spiliotis, E., Assimakopoulos, V., 2021. The M5 competition: Background, organization, and implementation. International Journal of Forecasting (In Press).
URL `https://doi.org/10.1016/j.ijforecast.2021.07.007`

Makridakis, S., Spiliotis, E., Assimakopoulos, V., 2022. The M5 Accuracy Competition: Results, Findings and Conclusions.
URL `https://doi.org/10.1016/j.ijforecast.2021.11.013`

McKenzie, E., 1986. Renormalization of seasonals in Winters' forecasting systems: Is it necessary? Operations Research 34 (1), 174–176.

Ord, K., Koehler, A. B., Snyder, R. D., 1997. Estimation and prediction for a class of dynamic nonlinear statistical models. Journal of the American Statistical Association 92 (440), 1621–1629.

Petropoulos, F., Svetunkov, I., 2020. A simple combination of univariate models. International Journal of Forecasting 36 (1), 110–115.

R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL `https://www.R-project.org/`

Rostami-Tabar, B., Babai, M. Z., Syntetos, A., Ducq, Y., sep 2013. Demand forecasting by temporal aggregation. Naval Research Logistics (NRL) 60 (6), 479–498.

Snyder, R. D., 1985. Recursive estimation of dynamic linear models. Journal of the Royal Statistical Society. Series B (Methodological) 47 (2), 272–276.

Snyder, R. D., Ord, J. K., Koehler, A. B., McLaren, K. R., Beaumont, A. N., 2017. Forecasting compositional time series: A state space approach. International Journal of Forecasting 33 (2), 502–512.

Snyder, R. D., Shami, R. G., 2001. Exponential smoothing of seasonal data: A comparison. Journal of Forecasting 20 (3), 197–202.

Spiliotis, E., Makridakis, S., Kaltsounis, A., Assimakopoulos, V., 2021. Product sales probabilistic forecasting: An empirical evaluation using the M5 competition data. International Journal of Production Economics 240 (June 2020), 108237.

Svetunkov, I., 2021a. greybox: Toolbox for Model Building and Forecasting. R package version 0.7.0.
URL `https://github.com/config-i1/greybox`

Svetunkov, I., 2021b. smooth: Forecasting Using State Space Models. R package version 3.1.2.
URL https://github.com/config-i1/smooth

Svetunkov, I., Boylan, J. E., 2020. State-space ARIMA for supply-chain forecasting. International Journal of Production Research 58 (3), 818–827.

Taylor, J. W., Snyder, R. D., 2012. Forecasting intraday time series with multiple seasonal cycles using parsimonious seasonal exponential smoothing. Omega 40 (6), 748 – 757, special Issue on Forecasting in Management Science.

Trapero, J. R., Kourentzes, N., Martin, A., 2015. Short-term solar irradiation forecasting based on dynamic harmonic regression. Energy 84, 289–295.

Wang, J.-J., Wang, J.-Z., Zhang, Z.-G., Guo, S.-P., 2012. Stock index forecasting based on a hybrid model. Omega 40 (6), 758–766, special Issue on Forecasting in Management Science.

Zhu, K., Thonemann, U. W., 2004. An adaptive forecasting algorithm and inventory policy for products with short life cycles. Naval Research Logistics (NRL) 51 (5), 633–653.