Managmeent Science

Working Paper 2018:1

Complex Exponential Smoothing for Seasonal Time Series

Ivan Svetunkov, Nikolaos Kourentzes

The Department of Management Science
Lancaster University Management School
Lancaster LA1 4YX
UK

Complex Exponential Smoothing for Seasonal Time Series

Ivan Svetunkov[a,*], Nikolaos Kourentzes[a]

[a] *Lancaster University Management School*
Department of Management Science, Lancaster, LA1 4YX, UK

Abstract

The general seasonal Complex Exponential Smoothing (CES) is presented in this paper. CES is based on conventional exponential smoothing and the notion of the "information potential", an unobservable part of time series that the classical models do not consider. The proposed seasonal CES can capture known forms of seasonality, as well as new ones that are neither strictly additive nor multiplicative. In contrast to exponential smoothing, CES can capture both stationary and non-stationary processes, giving it greater modelling flexibility. In order to choose between the seasonal and non-seasonal CES a model selection procedure is discussed in the paper. An empirical evaluation of the performance of the model against ETS and ARIMA on simulated and real data is carried out. The findings suggest that CES simplifies substantially model selection, and as a result the forecasting process, while performing better than the benchmarks in terms of forecasting accuracy.

Keywords: Exponential smoothing, model selection, complex variables, seasonality, complex exponential smoothing, information potential

1. Introduction

Exponential smoothing (ETS) is one of the most popular family of models used both in research and practice. The variety of the exponential smoothing

*Correspondance: I Svetunkov, Department of Management Science, Lancaster University Management School, Lancaster, Lancashire, LA1 4YX, UK.

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

models is wide and allows modelling different types of time series components. Hyndman et al. (2008) present a taxonomy which leads to 30 models with different types of error, trend and seasonal components.

ETS is not free of modelling challenges. First, the large number of model forms introduces a selection problem. Although the model variety allows capturing different types of processes, at the same time it makes it difficult to select a correct one for a time series. This is usually addressed by using an information criterion, typically the Akaike Information Criterion (Hyndman et al., 2002), though Billah et al. (2006) showed that using other information criteria did not lead to significant differences in forecasting performance. However recent research showed that choosing a single most appropriate exponential smoothing model for a time series may not lead to the most accurate forecast. As a result various combination approaches have been proposed in the literature. For example, Kolassa (2011) investigated combination of the different ETS forecasts using Akaike weights with good results.

Second, it is assumed in ETS framework that any time series may be decomposed into level, trend and seasonal components, which in real life are arbitrary and unobservable. For example, it may not always be easy to identify whether a series exhibits a changing level or a trend (Hyndman et al., 2008). Similar complications are relevant to identifying the seasonal component and its nature: whether it is additive or multiplicative.

Third, the combination approaches highlight that there may be composite ETS forms that are not captured by the known models. The combinations described by Kolassa (2011) result in such non-customary trend and seasonality forms. Kourentzes et al. (2014) argue that full identification of trend and seasonality is not straightforward with conventional modelling, showing the benefits of using multiple levels of temporal aggregation to that purpose, demonstrating forecasting performance improvements. Kourentzes & Petropoulos (2015) proceed to show that this problem is more acute under the presence of extreme values, such as outliers due to promotional activities, where again a similar combination approach leads to more accurate forecasts. We argue that these composite forms of exponential smoothing perform well because the type of time series components, as well as the interaction between them, may be too restrictive under ETS for some time series.

In this paper we aim to overcome these problems by using a more general exponential smoothing formulation that has been proposed by Svetunkov & Kourentzes (2015). It assumes that any time series can be modelled as the observed series value and an unobserved information potential. This leads to a less restrictive model without an arbitrary decomposition of time series. The authors implement this idea using complex variables, proposing the Complex Exponential Smoothing (CES). Their investigation is limited to non-seasonal time series, where they find that CES can accurately model time series without arbitrarily separating them into level and trend ones. Another crucial advantage of CES over conventional ETS is that it can capture both stationary and non-stationary processes.

Here we extend CES for seasonal time series, which leads to a a family of CES models that can model all types of level, trend, seasonal and trend seasonal time series in the conventional ETS classification. In contrast to ETS, CES has only two forms (non-seasonal and seasonal) and hence we deal with a simpler model selection problem that allows capturing different types of components and interactions between them. To test if the extended CES is capable of modelling appropriate time series structures, we conduct a simulation study and find that the simplified selection problem leads to better results in comparison with conventional ETS. We then evaluate the forecasting performance of extended CES against established benchmarks and find it to produce more accurate forecasts on the monthly M3 dataset. We argue that even though the formulation of CES appears to be more complicated than individual ETS models, the substantially simplified selection problem and its good accuracy makes it appealing for practice.

The rest of the paper is organised as follows: section 2 introduces the Complex Exponential Smoothing model and its seasonal extension. Section 3 presents the setup and provides the results of empirical evaluations on simulated and real data. This section is then followed by concluding remarks.

2. Complex Exponential Smoothing

2.1. Information Potential

A fundamental idea behind CES is the *Information Potential* (Svetunkov & Kourentzes, 2015). Any measured time series contains some information, which may be less than the time series in all its totality and potentially unobservable due to sampling. For example, that might be due to some unobserved long memory process or other more exotic structures.

It was shown that there is a convenient way to write the measured series and the information potential using a complex variable: $y_t + ip_t$ (Svetunkov, 2012), where y_t is the actual value of the series, p_t is the information potential on the observation t and i is the imaginary unit, the number that satisfies the equation: $i^2 = -1$. Svetunkov & Kourentzes (2015) proposed using to use the value of the error term as a proxy of information potential. This leads to desirable properties of CES that are discussed below.

2.2. Complex Exponential Smoothing

A complex counterpart to conventional exponential smoothing can be developed as:

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

where $\hat{y}_t$ is the forecast of the actual series, $\hat{p}_t$ is the estimate of the information potential, $\alpha_0 + i\alpha_1$ is complex smoothing parameter. Using the aforementioned proxy $p_t = \epsilon_t$, Svetunkov & Kourentzes (2015) derived the state-space model of CES that can be used to further explore its properties. It can be written the following way:

$$\begin{cases} y_t = l_{t-1} + \epsilon_t \\ l_t = l_{t-1} - (1-\alpha_1)c_{t-1} + (\alpha_0 - \alpha_1)\epsilon_t \\ c_t = l_{t-1} + (1-\alpha_0)c_{t-1} + (\alpha_0 + \alpha_1)\epsilon_t \end{cases}, \quad (2)$$

where l_t is the level component, c_t is the information component on observation t, and $\epsilon_t \sim N(0, \sigma^2)$. This state-space form of CES can be written in a more general form:

$$\begin{cases} y_t = w' x_{t-1} + \epsilon_t \\ x_t = F x_{t-1} + g\epsilon_t \end{cases}, \quad (3)$$

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

One of the main features of CES is that it does not contain an explicit trend: its l_t and c_t components are connected with each other and change in time depending on the value of the complex smoothing parameter. It can be shown that CES has an underlying ARIMA(2,0,2) model:

$$\begin{cases} (1 - \phi_1 B - \phi_2 B^2) y_t = (1 - \theta_{1,1} B - \theta_{1,2} B^2)\epsilon_t \\ (1 - \phi_1 B - \phi_2 B^2) \xi_t = (1 - \theta_{2,1} B - \theta_{2,2} B^2)\epsilon_t \end{cases}, \quad (4)$$

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 $\theta_{2,2} = \alpha_0 - \alpha_1 - 2$ and $\xi_t = p_t - c_{t-1}$ is an

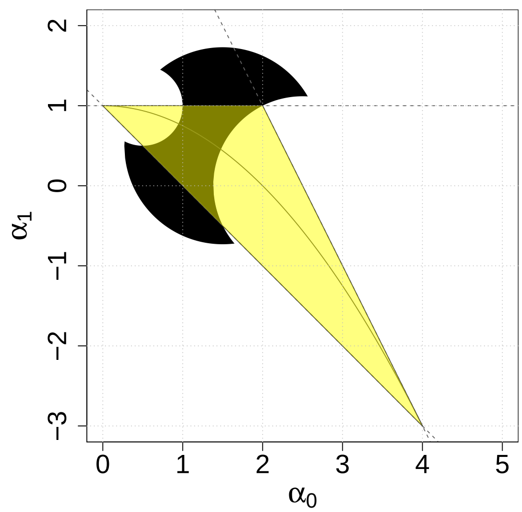


Figure 1: CES stability (the black area) and stationarity (the light triangle) conditions.

information gap, the value showing the amount of the information missing in the information component c_t.

However the parameter space for the autoregressive terms in this model differs from the conventional AR(2) model due to the connection of AR and MA terms via the complex smoothing parameter. It should be noted that the stationary condition is not essential for CES, which means that the roots of the characteristic equation in (4) may lie inside the unit circle. This gives the model an additional flexibility and allows to slide between level and trend time series rather than switch between them. This means in its turn that CES can be both stationary and non-stationary, depending on the complex smoothing parameter value, while all the conventional ETS models are strictly non-stationary.

Stability condition	Stationarity condition
$(\alpha_0 - 2.5)^2 + \alpha_1^2 > 1.25$	$\alpha_1 < 5 - 2\alpha_0$
$(\alpha_0 - 0.5)^2 + (\alpha_1 - 1)^2 > 0.25$	$\alpha_1 < 1$
$(\alpha_0 - 1.5)^2 + (\alpha_1 - 0.5)^2 < 1.5$	$\alpha_1 > 1 - \alpha_0$

Table 1: Stability and stationarity conditions of CES.

The stability and stationarity conditions for CES on the plane of complex

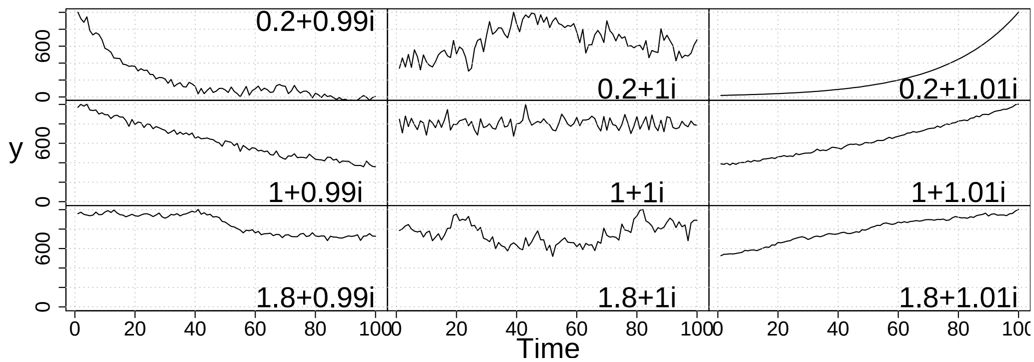


Figure 2: Data simulated using CES with different complex smoothing parameter values.

smoothing parameters are shown on the Figure 1. They correspond to the inequalities shown in the table 1. Note that while we restrict CES parameters with stability region, we do not impose similar restriction for stationarity allowing the model to capture and produce long-term trends.

CES can produce different types of trajectories depending on the complex smoothing parameter value. Some example trajectories are shown in the Figure 2. Note that the imaginary part of the complex smoothing parameter determines the direction of the trend (or the absence of it when $\alpha_1 = 1$), while the real part influences the steepness of trend. When $\alpha_0 < 1$ the trend has obvious exponential character. When $\alpha_0 = 1$ the trend slows down but still shows the features of the exponential function. Finally, when $\alpha_0 > 1$ the series reveals features of additive trend with possible change of level and slope. One of the interesting findings is that when $\alpha_0 + i\alpha_1 = 1 + i$ the generated series becomes stationary. The other feature of CES is that the behaviour of the trend with $\alpha_1 < 1$ resembles the damped trend model. Finally, if for some reason a level of series becomes negative, then the same complex smoothing parameter values will result in the opposite trajectories direction (for example, $\alpha_0 < 1$ will result in the increase either than decline).

We should also point out that CES is unable to produce trajectories with changing signs of long-term trends. The reason for this is because the imaginary part of complex exponential smoothing defines the long-term direction of the trend in the data. This can be considered as a drawback when an analyst is expecting the radical change of long-term trend in data. But this is also a benefit of the model, because it allows capturing long-term dependencies clearer than many other commonly used models do and does not react rapidly to possible changes in short-term trends.

This example shows that CES is capable of capturing different types of trends, as well as level series (in an ETS context). Svetunkov & Kourentzes showed empirically that a single CES model can be used instead of several ETS models with different types of trends for common forecasting tasks. This is one of the major advantages of the model. When disregarding seasonal time series no model selection is needed for CES, in contrast to conventional exponential smoothing.

2.3. Seasonal CES

Here we extend the CES model presented above to cater for seasonality. The simplest way to derive a seasonal model using CES is to take values of level and information components with a lag m which corresponds to the seasonality lag instead of $t-1$:

$$\begin{cases} y_t = l_{1,t-m} + \epsilon_t \\ l_{1,t} = l_{1,t-m} - (1-\beta_1)c_{1,t-m} + (\beta_0 - \beta_1)\epsilon_t \\ c_{1,t} = l_{1,t-m} + (1-\beta_0)c_{1,t-m} + (\beta_0 + \beta_1)\epsilon_t \end{cases} \quad (5)$$

where $l_{1,t}$ is the level component, $c_{1,t}$ is the information component on observation t, β_0 and β_1 are smoothing parameters. This formulation follows similar ideas to the reduced seasonal exponential smoothing forms by Snyder & Shami (2001) and Ord & Fildes (2012).

The model (5) preserves all the features of the original CES (2) and produces all the possible CES trajectories with the same complex smoothing parameter values. These trajectories, however, will be lagged and will appear for each seasonal element instead of each observation. This means that the model (5) produces non-linear seasonality which can approximate both additive and multiplicative seasonality depending on the complex smoothing parameter and initial components values. Note that the seasonality produced by the model may even have these features regardless of the value of the level, thus generating seasonalities that cannot be classified as either additive or multiplicative. Naturally, no known exponential smoothing model can produce similar patterns: multiplicative seasonality implies that the amplitude of fluctuations increases with the increase of the level, while with seasonal CES the amplitude may increase without it.

However this model is not flexible enough: it demonstrates all the variety of possible seasonality changes only when the level of the series is equal to zero (meaning that a part of the data lies in the positive and another lies in the negative plane). To overcome this limitation we extend the original model (2) with the basic seasonal model (5) in one general seasonal CES model:

$$\begin{cases} y_t = l_{0,t-1} + l_{1,t-m} + \epsilon_t \\ l_{0,t} = l_{0,t-1} - (1-\alpha_1)c_{0,t-1} + (\alpha_0 - \alpha_1)\epsilon_t \\ c_{0,t} = l_{0,t-1} + (1-\alpha_0)c_{0,t-1} + (\alpha_0 + \alpha_1)\epsilon_t \\ l_{1,t} = l_{1,t-m} - (1-\beta_1)c_{1,t-m} + (\beta_0 - \beta_1)\epsilon_t \\ c_{1,t} = l_{1,t-m} + (1-\beta_0)c_{1,t-m} + (\beta_0 + \beta_1)\epsilon_t \end{cases} \tag{6}$$

The model (6) still can be written in a conventional state-space form (2). It exhibits several differences from the conventional smoothing seasonal models. First, the proposed seasonal CES in (6) 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 some estimates of level and information components in the past and have more common features with seasonal ARIMA (Box & Jenkins, 1976, p.300) than with the conventional seasonal exponential smoothing models. Second, it can be shown that the general seasonal CES has an underlying model that corresponds to SARIMA$(2,0,2m+2)(2,0,0)_m$ (see Appendix A), which can be either stationary or not, depending on the complex smoothing parameters values.

The general seasonal CES preserves the properties of both models (2) and (5): it can produce non-linear seasonality and all the possible types of trends discussed above, as now the original level component $l_{0,t}$ can become negative while the lagged level component $l_{1,t}$ may become strictly positive.

Finally this model retains the interesting property of independence of the original level and lagged level components, so a multiplicative (or other) shape seasonality may appear in the data even when the level of the series does not change. This could happen for example when the seasonality is either nonlinear or some other variable is determining its evolution, as for example is the case with solar power generation (Trapero et al., 2015).

2.4. Parameters estimation

During the estimation of the parameters of the general seasonal CES some constrains should be introduced to achieve the stability of the model. The state-space model is stable when all the eigenvalues of the discount matrix $D = F - gw'$ lie inside the unit circle. Unfortunately, it is hard to derive the exact regions for the smoothing parameters of general seasonal CES but the stability condition can be checked during the optimisation. To do that the eigenvalues of the following discount matrix should be calculated (see Appendix B):

$$D = \begin{pmatrix} 1-\alpha_0+\alpha_1 & \alpha_1-1 & \alpha_1-\alpha_0 & 0 \\ 1-\alpha_0-\alpha_1 & 1-\alpha_0 & -\alpha_1-\alpha_0 & 0 \\ \beta_1-\beta_0 & 0 & 1-\beta_0+\beta_1 & \beta_1-1 \\ -\beta_1-\beta_0 & 0 & 1-\beta_0-\beta_1 & 1-\beta_0 \end{pmatrix} \tag{7}$$

The estimation of the parameters of CES can be done using the likelihood function. This is possible due to the additive form of the error term in (6). The likelihood function turns out to be similar to the one used in ETS models (Hyndman et al., 2002):

$$L(g, x_0, \sigma^2|y) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^T \exp\left(-\frac{1}{2}\sum_{t=1}^{T}\left(\frac{\epsilon_t}{\sigma}\right)^2\right) \tag{8}$$

This likelihood function value can be used in calculation of information criteria. The double negative logarithm of (8) can be calculated to that purpose:

$$-2\log(L(g, x_0|y)) = T\left(\log(2\pi e) + \log\left(\sum_{t=1}^{T}\epsilon_t^2\right)\right). \tag{9}$$

2.5. Model selection

Using (9) the Akaike Information Criterion for both seasonal and non-seasonal CES models can be calculated:

$$AIC = 2k - 2\log(L(g, x_0|y)) \tag{10}$$

where k is the number of coefficients and initial states of CES. For the non-seasonal model (2) k is equal to 4 (2 complex smoothing parameters and 2 initial states). For the seasonal model the number of the coefficients in (6) becomes much greater than in the original model: $k = 4 + 2m + 2$, which is 4 smoothing parameters, $2m$ initial lagged values and 2 initial values of the generic level and information components. Naturally, other information criteria can be constructed.

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 under conventional ETS.

3. Empirical evaluation

3.1. Model selection

To evaluate the performance of the model selection procedure we simulate series with known structure and attempt to model them with CES. Each series is simulated using ETS at a monthly frequency and contains 120 observations. All the smoothing parameters are generated using uniform distribution, restricted by the traditional bounds: $\alpha \in (0, 1), \beta \in (0, \alpha), \gamma \in (0, 1-\alpha)$. Normal distribution with zero mean is used in the error term generation. We generate series with either additive or multiplicative errors, using error standard deviation equal to 50 and 0.05 respectively. The initial value of the level is set to 5000, while the initial values of trend is set to 0 for the additive cases and to 1 for the multiplicative cases. This allows the model used as DGP to produce either growth or decline, depending on the error term and smoothing parameter values. All the initial values of seasonal components are randomly generated and then normalised. For each of the 9 process shown in Table 2 1000 time series are generated.

We fit the two types of CES and chose the one with the smallest AIC corrected for small sample sizes (AICc) value. As benchmarks we fit ETS and ARIMA (both use AICc for model selection). The implementation of CES was done in R and is available as the "CES" package (in github: `https://github.com/config-i1/CES`). The benchmarks are produced using the "forecast" package for R (Hyndman & Khandakar, 2008).

The percentage of the appropriate models chosen by CES, ETS and ARIMA are shown in the Table 2. The values in the table are the percentage of successful time series characteristics identified by each model. We note here that the definition of success of time series characteristics identification by different models is subjective. This especially concerns ARIMA model, for which we have developed a set of criteria (see explanation below), allowing to define if it captured any type of trend or seasonality in different situations. However the main objective of this experiment is to see whether CES is able to make a distinction between seasonal and non-seasonal data and compare performance of the model in this task with other models in controlled environment.

Column "CES" shows in how many instances the appropriate seasonal or non-seasonal model is chosen. We can observe that CES is very accurate and managed to select the appropriate model in the majority of cases. It is important to note that the fitted CES is not equivalent to the data generating

DGP	CES	ETS			ARIMA		
		Trend	Seasonal	Exact	Trend	Seasonal	Overall
$N(5000, 50^2)$	99.9	97.3	99.8	97.1	96.5	45.7	44.1
ETS(ANN)	99.1	88.0	99.7	49.3	51.5	46.5	28.0
ETS(MNN)	99.3	85.1	99.6	50.9	59.7	47.3	30.0
ETS(AAN)	91.5	94.4	99.7	82.3	96.4	45.7	43.5
ETS(MMN)	98.9	91.6	99.7	68.9	92.3	35.2	32.2
ETS(ANA)	100	85.4	100	46.3	53.0	100	53.0
ETS(AAA)	100	92.1	100	79.1	86.3	100	86.3
ETS(MNM)	100	65.9	100	32.6	61.9	100	61.9
ETS(MMM)	98.2	88.4	100	52.7	70.3	100	70.3
Average	98.5	87.6	99.8	62.1	74.2	68.9	49.9

Table 2: The percentage of the forecasting models chosen appropriately for each data generating process (DGP).

process in each time series, but nevertheless it is able to approximate the time series structure.

The column "ETS, Trend" shows in how many cases the trend or its absence is identified appropriately (not taking into account the type of trend). The "ETS, Seasonal" column shows similar accuracy in the identification of seasonal components. The lowest trends identification ETS accuracy is in the case of ETS(MNM) process with 65.9%, while the average accuracy in capturing trends appropriately is 87.6%. The accuracy of seasonal components identification in ETS is much better, with the average accuracy for all the DGPs in this case being 99.8%. The column "Exact" shows in how many cases the exact DGP is identified by ETS. The average accuracy of ETS in this task is 62.1%. Note that contrary to CES, ETS should have identified the exact model in the majority of cases, since it was used to generate the series. The reason behind this misidentification may be due to the information criterion used and the sample size of the generated time series.

Table 2 also provides results for the ARIMA benchmark. The column "ARIMA, Trend" shows the number of cases where ARIMA identifies the trend or its absence appropriately. Although trend in ARIMA is not defined directly, we use three criteria to indicate that ARIMA captured a trend in the time series: ARIMA has a drift, or ARIMA has second differences, or ARIMA has first differences with non-zero AR element. It can be noted that the lowest accuracy of ARIMA in trends identification is in the cases of level ETS models DGPs. The average accuracy of ARIMA in this task is 74.2%.

The column "ARIMA, Seasonal" shows the number of cases where ARIMA identifies the seasonality or its absence appropriately. The cases where either seasonal AR or seasonal MA or seasonal differences contains non-zero values are identified as seasonal ARIMA models. The values in the Table 2 indicate that ARIMA makes a lot of mistakes in identifying the seasonal models in time series generated using non-seasonal ETS models, which leads to the average accuracy of 68.9% in this category. Even though the generating processes is different than the model used in this case, capturing seasonality in the pure non-seasonal data makes no sense. The column "ARIMA, Overall" shows the number of cases where both trend and seasonality are identified appropriately by ARIMA using the criteria described above. Note in the "Overall" column we only count the cases where ARIMA appropriately identifies both the presence of trend and season, so as to make it comparable with the CES results. ARIMA identifies the appropriate form mainly in the cases of ETS(AAA) and ETS(MMM) models.

Obviously the ETS and ARIMA results are dependent on the identification methodology employed. But the comparison done here retains its significance since all three alternative models, CES, ETS and ARIMA use the same information criterion to identify the final model form.

This simulation experiment shows that CES, having only two models to choose from, becomes more efficient in time series identification than ETS and ARIMA models, that both have to choose from multiple alternatives. We argue that CES is capable of capturing both level and trend series (Svetunkov & Kourentzes, 2015) and its seasonal extension can produce both additive and multiplicative seasonality. This reduces the number of alternative models substantially. Furthermore any seasonal components are highly penalized with CES during model selection. Therefore the seasonal model is chosen only in cases when it fits data significantly better than its non-seasonal counterpart. This substantially simplifies the forecasting process as a whole.

3.2. Forecasting accuracy

The empirical out-of-sample accuracy evaluation is necessary to test the forecasting performance of CES and compare it against ETS and ARIMA. We conduct such an an evaluation on the monthly data from M3 Competition (Makridakis & Hibon, 2000) that contains both trend and seasonal time

	CES	ETS	ARIMA
Minimum	0.134	**0.084**	0.098
25% quantile	0.665	**0.664**	0.703
Median	**1.049**	1.058	1.093
75% quantile	**2.178**	2.318	2.224
Maximum	**28.440**	53.330	59.343
Mean	**1.922**	1.934	1.967

Table 3: MASE values of competing methods. The smallest values are in bold.

series. The forecasting horizon (18 periods ahead) is retained the same as in the original competition, however a rolling origin evaluation scheme is used, with the last 24 observations withheld.

The Mean Absolute Scaled Error (MASE) is used to compare the performance of models for each forecast horizon from each origin (Hyndman & Koehler, 2006):

$$MASE = \frac{\sum_{j=1}^{h} |e_{T+j}|}{\frac{h}{n-1} \sum_{t=2}^{T} |y_t - y_{t-1}|} \tag{11}$$

Using these errors we estimate quantiles of the distribution of MASE along with the mean values from each origin. Table 3 presents the results of this experiment.

CES demonstrates the smallest mean and median MASE. It also has the smallest maximum error which indicates that when all the models failed to forecast some time series accurately, CES was still closer to the actual values. In the contrast ETS had the smallest minimum MASE, while CES had the highest respective value among all the competitors. This means that CES may not be as accurate as other models in the cases when the time series is relatively easy.

To see if the difference in the forecasting accuracy between CES and the other methods is significant, we use the Multiple Comparisons with Best (MCB) test (Koning et al., 2005). We note here that (Hibon et al., 2012) demonstrated that MCB is a special case of Nemenyi test. The results of this test indicate that CES is significantly more accurate than ETS and ARIMA (see Figure 3).

To investigate what could cause this difference in forecasting accuracy, the

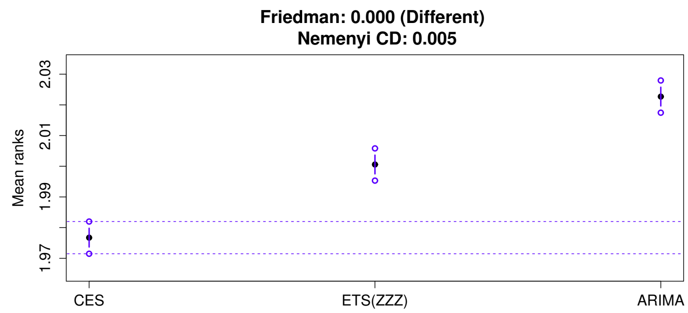


Figure 3: The results of MCB test on monthly data of M3. The dotted lines are the critical distances for the best model and we can see that both ETS and ARIMA are found to be significantly different.

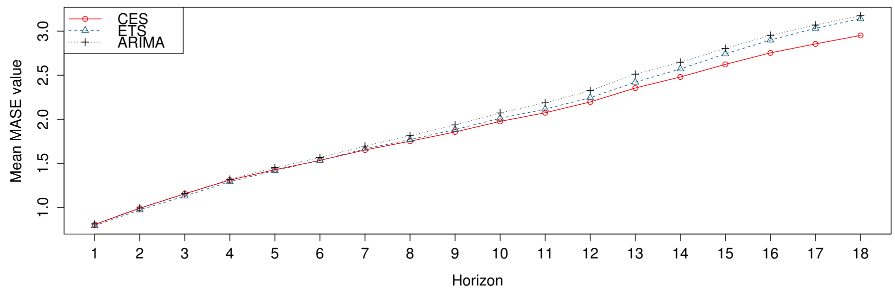
mean and median MASE are calculated for each forecast horizon, obtaining the one-step ahead, two-steps ahead etc. mean and median MASE values. These values are shown on the Figures 4a and 4b.

The analysis of the Figure 4 shows that while the errors of the methods are very similar for short horizons (with ETS being slightly better), the difference between them increases on longer horizons. The difference in mean values of MASE between methods starts increasing approximately after a year (12 observations) with CES being the most accurate. The same feature is present in median MASE values.

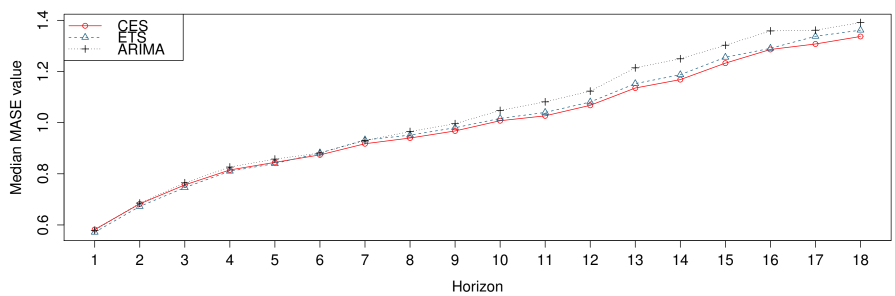
Concluding the results of this experiment, CES performs significantly better than ETS and ARIMA on the monthly M3 data. CES demonstrates that it can model various types of time series successfully and is particularly accurate in relation to the benchmark models for longer horizons. We attribute this to the way that time series trends and seasonal components are modelled under CES.

4. Conclusions

We proposed a new type of seasonal exponential smoothing, based on Complex Exponential Smoothing and discussed the model selection procedure that can be used to identify whether the non-seasonal or seasonal CES should be selected for time series forecasting. While the non-seasonal CES can produce different non-linear trajectories and approximate additive and multiplicative trends, the seasonal model produces non-linear seasonality



(a) Mean MASE



(b) Median MASE

Figure 4: Mean and median MASE value for CES, ETS and ARIMA on different horizons.

that can approximate both additive and multiplicative seasonal time series and even model new forms of seasonality that do not strictly lie under either of these cases. The latter can be achieved as a result of the independence of the seasonal and non-seasonal components and non-linearity of CES.

We also discussed the statistical properties of the general seasonal CES and showed that it has an underlying SARIMA$(2, 0, 2m+2), (2, 0, 0)_m$ model. This model can be either stationary or not, depending on complex smoothing parameter values, which gives CES its flexibility and a clear advantage over the conventional exponential smoothing models which are strictly non-stationary.

The simulation conducted showed that the proposed model selection procedure allows choosing the appropriate CES effectively, making only few mistakes in distinguishing seasonal from non-seasonal time series. In comparison ETS and ARIMA were not as effective as CES in this task. We argue that this is a particularly useful feature of the CES model family. Although the models initially may appear more complicated than conventional ETS, the fact that a forecaster needs to consider only two CES variants, while being able to model a wide variety of trend and seasonal shapes, can greatly simplify the forecasting process.

The forecast competition between CES, ETS and ARIMA conducted on the monthly data from M3 showed that CES achieved both the lowest mean and median forecasting errors. The MCB test showed that the differences in forecasting accuracy between these models were statistically significant. Further analysis showed that the superior performance of CES was mainly caused by the more accurate long-term forecasts. This is attributed to the ability of CES to capture the long-term dependencies in time series, due to its non-linear character that permits CES to obtain more flexible weight distributions in time than ETS.

Overall CES proved to be a good model that simplifies the model selection procedure substantially and allows increasing the forecasting accuracy.

We should note that we have used error term as a proxy in CES framework, but there can be suggested other proxies instead. For example, using differences of data as proxy instead of error term leads to underlying ARMA(3,2) model and introduces other properties of Complex Exponential Smoothing. However study of these properties is out of the scope of this paper.

Finally, it would be interesting to compare CES with models with long memory, such as ARARMA and ARFIMA. Such future research would permit to highlight similarities and differences between these models, but also provide additional evidence of the forecasting performance of such models, which as is the case for ARARMA have been potentially not explored to their potential.

Acknowledgements

We would like to thank the two anonymous reviewers and the associate editor for their constructive communication that has helped improve the paper.

Appendix A. General seasonal CES and SARIMA

The model (6) can be written in the following state-space form:

$$\begin{aligned} y_t &= w_0' x_{0,t-1} + w_1' x_{1,t-m} + \epsilon_t \\ x_{0,t} &= F_0 x_{0,t-1} + g_0 \epsilon_t \\ x_{1,t} &= F_1 x_{1,t-m} + g_1 \epsilon_t \end{aligned} , \tag{A.1}$$

where $x_{0,t} = \begin{pmatrix} l_{0,t} \\ c_{0,t} \end{pmatrix}$ is the state vector of the non-seasonal part of CES, $x_{1,t} = \begin{pmatrix} l_{1,t} \\ c_{1,t} \end{pmatrix}$ is the state vector of the seasonal part, $w_0 = w_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ are the measurement vectors, $F_0 = \begin{pmatrix} 1 & \alpha_1 - 1 \\ 1 & 1 - \alpha_0 \end{pmatrix}$, $F_1 = \begin{pmatrix} 1 & \beta_1 - 1 \\ 1 & 1 - \beta_0 \end{pmatrix}$ are transition matrices and $g_0 = \begin{pmatrix} \alpha_1 - \alpha_0 \\ \alpha_1 + \alpha_0 \end{pmatrix}$, $g_1 = \begin{pmatrix} \beta_1 - \beta_0 \\ \beta_1 + \beta_0 \end{pmatrix}$ are persistence vectors of non-seasonal and seasonal parts respectively.

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

$$\begin{aligned} y_t &= w' x_{t-l} + \epsilon_t \\ x_t &= F x_{t-l} + g \epsilon_t \end{aligned} , \tag{A.2}$$

where $x_t = \begin{pmatrix} x_{0,t} \\ x_{1,t} \end{pmatrix}$, $x_{t-l} = \begin{pmatrix} x_{0,t-1} \\ x_{1,t-m} \end{pmatrix}$, $w = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix}$, $F = \begin{pmatrix} F_0 & 0 \\ 0 & F_1 \end{pmatrix}$, $g = \begin{pmatrix} g_0 \\ g_1 \end{pmatrix}$. The state vector x_{t-l} can also be rewritten as $x_{t-l} = \begin{pmatrix} B & 0 \\ 0 & B^m \end{pmatrix} \begin{pmatrix} x_{0,t} \\ x_{1,t} \end{pmatrix}$, where B is a backshift operator. Making this substitution and taking $L = \begin{pmatrix} B & 0 \\ 0 & B^m \end{pmatrix}$ the state-space model (A.2) can be transformed into:

$$\begin{aligned} y_t &= w' L x_t + \epsilon_t \\ x_t &= F L x_t + g \epsilon_t \end{aligned} \tag{A.3}$$

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

$$(I_2 - FL) x_t = g \epsilon_t, \tag{A.4}$$

which after a simple manipulation leads to:

$$x_t = (I_2 - FL)^{-1} g \epsilon_t, \tag{A.5}$$

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

$$y_t = w' L (I_2 - FL)^{-1} g \epsilon_t + \epsilon_t. \tag{A.6}$$

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

$$y_t = (1 + w_0'(I_2 - F_0 B)^{-1} g_0 B + w_1'(I_2 - F_1 B^m)^{-1} g_1 B^m)\epsilon_t. \tag{A.7}$$

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

$$\begin{aligned} y_t = &\left(1 + w_0' \begin{pmatrix} 1-B & (1-\alpha_1)B \\ -B & 1-B+\alpha_0 B \end{pmatrix}^{-1} \begin{pmatrix} \alpha_1 - \alpha_0 \\ \alpha_1 + \alpha_0 \end{pmatrix} B + \right. \\ &\left. w_1' \begin{pmatrix} 1-B^m & (1-\beta_1)B^m \\ -B^m & 1-B^m+\beta_0 B^m \end{pmatrix}^{-1} \begin{pmatrix} \beta_1 - \beta_0 \\ \beta_1 + \beta_0 \end{pmatrix} B^m)\epsilon_t \right). \end{aligned} \tag{A.8}$$

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

$$(I_2 - F_0 B)^{-1} = \frac{1}{1 - 2B - (\alpha_0 + \alpha_1 - 2)B^2} \begin{pmatrix} 1-(1-\alpha_0)B & (\alpha_1 - 1)B \\ B & 1-B \end{pmatrix}, \tag{A.9}$$

similarly the inverse of the second matrix is:

$$(I_2 - F_1 B^m)^{-1} = \frac{1}{1 - 2B^m - (\beta_0 + \beta_1 - 2)B^{2m}} \begin{pmatrix} 1-(1-\beta_0)B^m & (\beta_1 - 1)B^m \\ B^m & 1-B^m \end{pmatrix}. \tag{A.10}$$

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

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

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

Appendix B. Discount matrix of the general seasonal CES

Substituting the error term in the transition equation (A.2) by the value from the measurement equation leads to:

$$x_t = Fx_{t-l} + gy_t - gw'x_{t-l} = (F - gw')x_{t-l} + gy_t, \tag{B.1}$$

which leads to the following discount matrix:

$$D = F - gw' = \begin{pmatrix} 1-\alpha_0+\alpha_1 & \alpha_1-1 & \alpha_1-\alpha_0 & 0 \\ 1-\alpha_0-\alpha_1 & 1-\alpha_0 & -\alpha_1-\alpha_0 & 0 \\ \beta_1-\beta_0 & 0 & 1-\beta_0+\beta_1 & \beta_1-1 \\ -\beta_1-\beta_0 & 0 & 1-\beta_0-\beta_1 & 1-\beta_0 \end{pmatrix} \tag{B.2}$$

References

Billah, B., King, M. L., Snyder, R. D., & Koehler, A. B. (2006). Exponential smoothing model selection for forecasting. *International Journal of Forecasting*, *22*, 239–247.

Box, G., & Jenkins, G. (1976). *Time Series Analysis, Forecasting And Control*. Holden-day, Oakland, California.

Hibon, M., Crone, S. F., & Kourentzes, N. (2012). Statistical significance of forecasting methods an empirical evaluation of the robustness and interpretability of the mcb, anom and friedman-nemenyi test. *The proceedings of the 32nd Annual international Symposium on Forecasting*, *1*.

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

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*, 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*, 439–454.

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

Koning, A. J., Franses, P. H., Hibon, M., & Stekler, H. (2005). The m3 competition: Statistical tests of the results. *International Journal of Forecasting*, *21*, 397–409.

Kourentzes, N., & Petropoulos, F. (2015). Forecasting with multivariate temporal aggregation: The case of promotional modelling. *International Journal of Production Economics*, *In Press*.

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

Makridakis, S., & Hibon, M. (2000). The m3-competition: Results, conclusions and implications. *International Journal of Forecasting*, *16*, 451–476.

Ord, K., & Fildes, R. (2012). *Principles of business forecasting*. Cengage Learning.

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

Svetunkov, I., & Kourentzes, N. (2015). Complex exponential smoothing. *Working Paper of Department of Management Science, Lancaster University*, *2015:1*, 1–31.

Svetunkov, S. (2012). *Complex-Valued Modeling in Economics and Finance*. SpringerLink : Bücher. Springer.

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