Please cite this paper as:

Svetunkov Ivan et al. (2023). Incorporating Parameters Uncertainty in ETS. Working Paper of Department of Management Science, Lancaster University, 2023:2 , 1–19.

Management Science

Working Paper 2023:2

Incorporating Parameters Uncertainty in ETS

Ivan Svetunkov, Kandrika Pritularga

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

LUMS home page: https://lums.lancaster.ac.uk

Incorporating Parameters Uncertainty in ETS

Ivan Svetunkov[a,*], Kandrika Pritularga[a]

[a] *Centre for Marketing Analytics and Forecasting, Department of Management Science, Lancaster University Management School, UK*

Abstract

Exponential smoothing in ETS form has showed a great performance in terms of point forecasts in many competitions and practical applications. However, its capability of producing prediction intervals has not been investigated as thoroughly. In fact, literature shows that ETS tends to produce narrower than needed forecasts, underestimating the future uncertainty. In this paper, we look into one of the sources of the uncertainty that has been ignored by ETS, namely the uncertainty coming from the estimates of parameters, and develop a procedure that allows taking mitigating it. We show how the procedure can be used in practice and find that it indeed produces wider prediction intervals. Finally, we find that none of the existing approaches for prediction intervals generates the correct quantiles, and that the fundamental issue with all of them is that they are miscalibrated.

Keywords: forecasting, state-space model, estimation, prediction intervals

1. Introduction

Exponential smoothing is one of the popular dynamic models used in practice (Weller and Crone, 2012). It is robust, easy to work with and easy to explain to people who do not know statistics. Its statistical counterpart, the ETS model (Error-Trend-Seasonality) in the Single Source of Error (SSOE) framework has gained popularity in the academic community and amongst the practitioners with more statistical backgrounds (data scientists, business analysts and machine learning experts). It has performed well in many competitions, showing to be a robust model, in many cases outperforming or at least comparable to more complicated models (see, for example, Makridakis and Hibon, 2000; Athanasopoulos et al., 2011; Makridakis et al., 2020, 2022). Being well established, it is used in many papers as a benchmark model, giving hard time to machine learning community to beat in terms of point forecasts (see, for example, discussion in Kolassa, 2022).

However, when it comes to predictive distribution generated by ETS, it tends to underestimate the uncertainty, producing lower coverage than expected (Athanasopoulos et al., 2011).

*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)

The distribution is calculated analytically based on Hyndman et al. (2008, see derivations in Chapter 6), and it assumes that the parameters, e.g., smoothing parameters, initial values, and dampening parameters, are known. This assumption is sensible when we have a sufficient sample size to achieve consistent and efficient parameters asymptotically. However, in reality, the parameters can be biased and inefficient due to limited sample sizes. Note that this analytical predictive distribution is based on a closed-form formulae, and is available for the pure additive ETS models. There are some approximations for some mixed models and there are no analytical formulae for the models with the multiplicative trend.

In this paper, we propose a novel approach taking the uncertainty of estimates of parameters into account and propagating it to the prediction intervals for all ETS models, irrespective of the types and the number of components.

2. Literature review

One of the few papers that deals with the parameters uncertainty in ETS is Bergmeir, Hyndman, and Benítez (Bergmeir et al.), who proposed so called "Bagged ETS". The idea of their approach can be summarised in the following steps:

1. Decompose time series using STL decomposition (Seasonal Transformation using LOESS by Cleveland et al., 1990);
2. Extract and bootstrap the residuals;
3. Recreate time series based on STL and bootstrapped residuals;
4. Apply ETS to the new time series;
5. Produce point forecasts;
6. Repeat (2) - (5) many times;
7. Combine forecasts.

Their approach was focused on point forecasts, and the authors have shown that it leads to an increase in accuracy, although at the cost of computational time.

Petropoulos et al. (2018) explained why the Bagged ETS works and leads to the increase in accuracy. First, due to different samples generated in step (3), applied ETS will have different estimates of parameters, which mitigates the parameter uncertainty. Second, different ETS models might be applied to different subsamples, mitigating the model specification uncertainty. Thus the approach deals with two out of three fundamental sources of uncertainty outlined by Chatfield (1996).

However, there are several issues with the Bagged ETS. First, it assumes that STL decomposition is appropriate. While it is one of the most flexible decomposition approaches, it has its own assumptions, which might be violated. Furthermore, STL is an approach external to ETS, which also means that it is challenging to extend it to ETS with explanatory variables, for the multiple seasonal data or intermittent time series. Third, it is computationally intensive, requiring many iterations to obtain good results. Finally, it was designed to improve point forecasts, making it challenging to extend it to predictive distributions and generation of prediction intervals.

3. A case of pure additive ETS models

A case of pure additive ETS models is a special one, because it is possible to derive variance h-steps ahead conditional on the values available on the most recent observation. This variance can then be used to create prediction intervals, assuming, for example, normality of the residuals. The formulae derived by Hyndman et al. (2008) all assume that the values of all parameters are known. However, we show how the formulae can be derived in the following two scenarios:

1. When the values of the initial state vector are estimated (not known), while the smoothing parameters are known;
2. When the model parameters (e.g. smoothing parameters) are estimated, but the initial states are provided.

Note that we do not consider the case, where dampening parameter is estimated, because it is not possible to derive closed form of the conditional variance in that case.

In this section we will derive variances for the general pure additive state space model, which underlies all pure additive ETS and all ARIMA models (it is formulated similarly to how it was done in Svetunkov et al., 2023; Svetunkov, 2023a). We will focus on a non-seasonal case, for which all the lags of components are equal to one (which leads to the same model as in Hyndman et al., 2008):

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

where $\mathbf{w}$ is the measurement vector, $\mathbf{F}$ is the transition matrix, $\mathbf{g}$ is the persistence vector, $\mathbf{v}_{t-1}$ is the vector of lagged components.

3.1. Estimated initials

To capture the effect of the uncertainty of the initial state on the conditional h-steps ahead variance, we need to use recursions, discussed in Chapter 6 of Hyndman et al. (2008). The recursive relation from the first observation till some observation t can be written as:

$$\hat{\mathbf{v}}_t = \mathbf{D}^t \hat{\mathbf{v}}_0 + \sum_{j=0}^{t-1} \mathbf{D}^j \mathbf{g} y_{t-j}, \tag{2}$$

where $\mathbf{D} = \mathbf{F} - \mathbf{g}\mathbf{w}'$ is the discount matrix and the $\hat{}$ symbol denotes the estimated variable. The equation (2) demonstrates that the most recent value of the state vector depends on the linear combination of actual values and the discounted initial value $\hat{\mathbf{v}}_0$. Taking the variance of the most recent state conditional on all the previous actual values gives:

$$\mathrm{V}(\hat{\mathbf{v}}_t | y_1, y_2, \dots y_t) = \mathrm{V}(\hat{\mathbf{v}}_t | t) = \mathrm{V}\left(\mathbf{D}^t \hat{\mathbf{v}}_0\right) + \mathrm{V}\left(\sum_{j=0}^{t-1} \mathbf{D}^j y_{t-j} | t\right). \tag{3}$$

There is no covariance term between the first and the second elements in (2), because all the actual values in-sample are known and are not random. For the same reason, the second variance in (3) equals to zero as well leaving us with:

$$\mathrm{V}(\hat{\mathbf{v}}_t | t) = \mathbf{D}^t \mathrm{V}\left(\hat{\mathbf{v}}_0\right) \left(\mathbf{D}^t\right)', \tag{4}$$

where the symbol $'$ denotes transposition. The variance (4) can then be used for the calculation of the variance of the fitted value on observation t, given all the observations until t:

$$\mathrm{V}(\hat{y}_t | t) = \mathbf{w}' \mathbf{D}^{t-1} \mathrm{V}\left(\hat{\mathbf{v}}_0\right) \left(\mathbf{D}^{t-1}\right)' \mathbf{w}, \tag{5}$$

which in turn can be used in the calculation of the h-steps ahead variance conditional on all the information available until observation t:

$$\begin{aligned} \mathrm{V}(y_{t+h} | t) = & \mathbf{w}' \mathbf{F}^{h-1} \mathbf{D}^{t-1} \mathrm{V}\left(\hat{\mathbf{v}}_0\right) \left(\mathbf{D}^{t-1}\right)' \left(\mathbf{F}'\right)^{h-1} \mathbf{w} + \\ & \left(\left(\mathbf{w}' \sum_{j=1}^{h-1} \mathbf{F}^{j-1} \mathbf{g}\mathbf{g}' (\mathbf{F}')^{j-1} \mathbf{w} \right) + 1 \right) \sigma^2 \end{aligned}, \tag{6}$$

where σ^2 is the variance of the error term. To make this idea clearer, here how the h-steps ahead variance conditional on all the available information until t is calculated for the ETS(A,N,N)

model:

$$\begin{aligned} \mathrm{V}(y_{t+h}|t) =&(1-\alpha)^{t-1}\mathrm{V}\left(\hat{l}_0\right)(1-\alpha)^{t-1}+ \\ &\left(1+(h-1)\alpha^2\right)\sigma^2 \end{aligned} . \tag{7}$$

The formula (7) consists of two parts: the first element captures the uncertainty coming from the estimate of the initial value, while the second one captures the uncertainty coming from the smoothing parameter of the model, which in this specific case is assumed to be known. Note that the impact of the variance of the initial state on the conditional one decreases with the increase of t because of the exponentiation of the smoothing parameters in (6). If all the values of the discount matrix $\boldsymbol{D}$ are close to zero then the effect of the variance of the initial states will be negligible. In case of ETS(A,N,N) it corresponds to situation, when α is close to one.

3.2. Estimated smoothing parameters

To obtain the variance in case of estimated smoothing parameters, we need to refer to another recursion, which tracks how the state vector h-steps ahead depends on the most recent values and a linear combination of the residuals:

$$\mathbf{v}_{t+h-1} = \mathbf{F}^{h-1}\mathbf{v}_t + \sum_{j=1}^{h-1}\mathbf{F}^{j-1}\hat{\mathbf{g}}e_{t+h-j}, \tag{8}$$

where now the persistence vector is estimated, while the initial state $\mathbf{v}_0$ is assumed to be known. The variance of (8) is straightforward and can be written as:

$$\begin{aligned} \mathrm{V}(\mathbf{v}_{t+h-1}|t) =&\mathrm{V}\left(\sum_{j=1}^{h-1}\mathbf{F}^{j-1}\hat{\mathbf{g}}e_{t+h-j}|t\right) = \\ &\sum_{j=1}^{h-1}\mathrm{V}\left(\mathbf{F}^{j-1}\hat{\mathbf{g}}e_{t+h-j}|t\right) + 2\sum_{j=2}^{h-1}\sum_{i=1}^{j-1}\mathrm{cov}(\mathbf{F}^{j-1}\hat{\mathbf{g}}e_{t+h-j}, \mathbf{F}^{i}\hat{\mathbf{g}}e_{t+h-i}|t) \end{aligned} . \tag{9}$$

Each variance in (9) can be expressed as:

$$\mathrm{V}\left(\mathbf{F}^{j-1}\hat{\mathbf{g}}e_{t+h-j}|t\right) = \mathbf{F}^{j-1}\left(\mathrm{V}(\hat{\mathbf{g}}) + \mathrm{E}(\hat{\mathbf{g}})\mathrm{E}(\hat{\mathbf{g}})'\right)(\mathbf{F}^{j})'\sigma^2,$$

while the covariances will be equal to zero if the basic assumptions of statistical models hold (i.e. no autocorrelation of the residuals and expectation of residuals equals to zero). This means

that we get the following formula for the variance of the state:

$$\mathrm{V}(\mathbf{v}_{t+h-1}|t) = \sum_{j=1}^{h-1} \mathbf{F}^{j-1} \left(\mathrm{V}(\hat{\mathbf{g}}) + \mathrm{E}(\hat{\mathbf{g}})\mathrm{E}(\hat{\mathbf{g}})' \right) (\mathbf{F}^{j})'\sigma^2,$$

meaning that the final variance h-steps ahead conditional on all the available information until observation t is:

$$\mathrm{V}(y_{t+h}|t) = \left(\mathbf{w}' \sum_{j=1}^{h-1} \mathbf{F}^{j-1} \left(\mathrm{V}(\hat{\mathbf{g}}) + \mathrm{E}(\hat{\mathbf{g}})\mathrm{E}(\hat{\mathbf{g}})' \right) (\mathbf{F}^{j})'\mathbf{w} + 1 \right) \sigma^2. \tag{10}$$

for the same ETS(A,N,N), the conditional variance (10) becomes:

$$\mathrm{V}(y_{t+h}|t) = \left((h-1) \left(\mathrm{V}(\hat{\alpha}) + \hat{\alpha}^2 \right) + 1 \right) \sigma^2,$$

which now has not only the impact of the estimate of the smoothing parameter on the variance, but also the impact of its variance.

Unfortunately, it is not possible to derive formulae for the case of estimation of both smoothing parameters and the initial state, because then we introduce complex interactions between the vector $\hat{\mathbf{g}}$ and $\hat{\mathbf{v}}_t$. Still the formulae (6) and (10) are useful because they show how specifically the parameters and initial values uncertainty impacts the conditional h-steps ahead variance, which is typically used for prediction intervals construction.

4. Method of scenarios in ETS

The case of a pure additive ETS model discussed in the previous Section is useful, but quite restrictive. We need to have a more universal solution, allowing us propagating the uncertainty of initial states, smoothing and dampening parameters to the final prediction interval. We propose a Monte Carlo style approach, where different ETS models are applied to the data and are used for forecasting. It can be summarised in the following seven steps:

1. Generate random values for initial state and parameters of the estimated model;
2. Apply ETS model with the generated state and parameters;
3. Calculate the variance of the residuals;
4. Generate n random errors for the next h observations;
5. Use the final state on the observation t, the parameters and the generated errors to simulate future paths for $j = 1, \ldots, h$. Record the values;

6. Repeat steps (1) – (5) p times to record all the future paths;
7. Take quantiles of all the recorded future paths from (6).

The issue with this procedure is that it requires $p \times n$ simulated paths, which could be potentially computationally expensive. Another challenge is how to generate value in step (1). We argue that they can be generated from a multivariate Normal distribution with the mean $\boldsymbol{\theta}$ (the vector of all parameters) and the covariance matrix $\boldsymbol{\Sigma_\theta}$ (the covariance matrix of parameters). This should be possible to do if likelihood is used in the model estimation, which gives consistent and efficient estimates of parameters and should guarantee that the Central Limit Theorem holds, i.e. the estimates of parameters follow a multivariate Normal distribution: $\hat{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{\theta}, \boldsymbol{\Sigma_\theta})$. Note however that the smoothing and dampening parameters are typically restricted with some bounds (either the traditional from 0 to 1 or with admissible as in Chapter 10 of Hyndman et al., 2008). This means that instead of the Normal distribution they should follow the rectified Normal one (the one, for which all the values above the threshold are equal to the threshold).

While the estimate of $\boldsymbol{\theta}$ can be obtained when we optimise models parameters (e.g. via likelihood maximisation), obtaining the estimate of the covariance matrix of parameters $\boldsymbol{\Sigma_\theta}$ is more challenging. We suggest three options for this:

1. Hessian-based covariance matrix;
2. Bootstrap;
3. Heuristics.

The estimation method (1) relies on the numerical calculation of the Hessian matrix for parameters, then based on that – Fisher Information (which is the negative Hessian) and after that – the covariance matrix via:

$$\mathrm{V}(\hat{\boldsymbol{\theta}}) = \mathrm{FI}(\hat{\boldsymbol{\theta}})^{-1}, \tag{11}$$

where $\hat{\boldsymbol{\theta}}$ is the vector of the estimated parameters. This approach might be computationally expensive if the vector of parameters $\hat{\boldsymbol{\theta}}$ is large because in that case the numerical estimation of Hessian might be slow.

The alternative is the method (2), which has some specificity in case of dynamic (time series) models. We suggest using case resampling technique, according to which the sub-samples of the original data are taken, and the model is applied to them to get estimates of parameters. However, not to break the time series structure, we can only take a consecutive observations, picking randomly the start and the end points of the sample. Furthermore, if we are interested

in estimating the initial state vector, we cannot change the starting point – it needs to be one and the same point in time. This makes the bootstrap restrictive and can potentially lead to the underestimation of the uncertainty around the parameters. The method is computationally expensive where the complexity grows with the sample size, but should not change substantially with the increase of the vector of parameters $\hat{\boldsymbol{\theta}}$.

Finally, method (3) is to make the covariance matrix of parameters diagonal and to define variances as a percentage of the estimated value of parameter:

$$\mathrm{V}(\hat{\boldsymbol{\theta}}) = |\hat{\boldsymbol{\theta}}| \times \nu, \tag{12}$$

where ν is the proportion defined by the user (for example, $\nu = 0.1$ to set variance as a 10% of the value of the parameter). In general, this is a poor estimate of the true variance of parameters, but it is fast and could potentially work in some situations (we investigate them in the next sections).

The method described in this Section is universal and should work for any ETS model with any combination components and/or with explanatory variables. We have implemented it in the `reapply()` and `reforecast()` functions for the `smooth` package v3.2.2 (Svetunkov, 2023b) in R (R Core Team, 2023).

5. Real data experiment

5.1. The setup

To assess the performance of the proposed approach, we conduct an experiment on M1 (Makridakis et al., 1982), M3 (Makridakis and Hibon, 2000) and Tourism (Athanasopoulos et al., 2011) competitions data. These three datasets comprise overall 5316 time series of annual, quarterly, monthly and other frequency. The forecast horizon used for each time series is standard and comes from the competitions. We are aware of the newer competitions, but we consciously do not use their data because the quality of the M4 dataset has been criticised in the literature (Fildes, 2020; Ingel et al., 2020), while the M5 dataset is not relevant to our topic (it contains intermittent data).

We used `adam()` function from the `smooth` package in R to estimate ETS models and then used `forecast()` and `reforecast()` functions to generate several types of prediction intervals:

1. Parametric – intervals generated using the conventional approach to ETS;

2. Semiparametric – intervals generated based on Lee and Scholtes (2014) method, relying on multistep forecast errors, produced in-sample;
3. Nonparametric – Taylor and Bunn (1999) method, modified to use power function instead of the polynomial (see Subsection 18.3.4 of Svetunkov, 2023a);
4. Empirical – taking quantiles of the multistep forecast errors, produced in-sample, as discussed in Lee and Scholtes (2014);
5. ETS Forecast – parametric prediction intervals generated by `ets()` function from the `forecast` package in R (Hyndman and Khandakar, 2008), needed for benchmarking only;
6. Complete – the Hessian-based approach discussed in the Section 4;
7. Complete bootstrap, the second method discussed in Section 4;
8. Complete heuristics with 1%, 5%, 10%, 15% and 20% – the heuristics approach, also discussed in Section 4.

We generated intervals with levels from 2% to 98% to get quantiles from 1% to 99%. We then used three measures to summarise performance of the approaches:

1. Coverage – showing the percentage of values lying inside the prediction interval. The ideal coverage should coincide with the nominal one (e.g. having 95% of actual holdout values lying inside the prediction interval with 95% confidence level). We use the coverage bias by subtracting the target one from the achieved one. As a result, in the ideal situation the number we obtain should all be equal to zero. If they are positive, this means that our intervals were wider than needed.
2. Quantile coverage – showing, how many observations lie below the nominal level. This measure is similar to the 1, but instead of measuring values between bounds, looks at those that lie below it.
3. Scaled range – showing the spread of the intervals. The lower it is, the narrower the interval is;

For simplicity, the scaling of (3) is done by dividing the values by the mean in-sample actual value, similar to how it was done in Petropoulos and Kourentzes (2015).

5.2. The results

Summarising the results of the experiment, we excluded 77 time series, because they required manual tuning (contained outliers, which impacted the performance of ETS models).

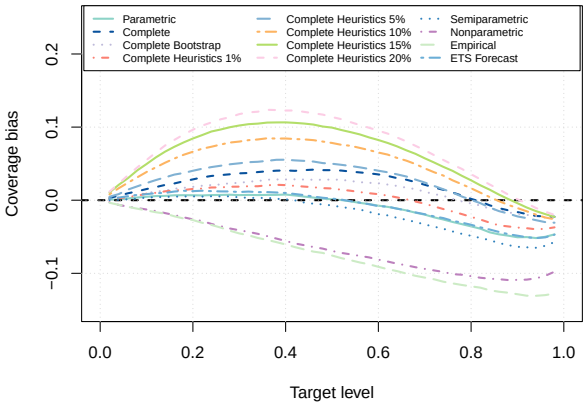


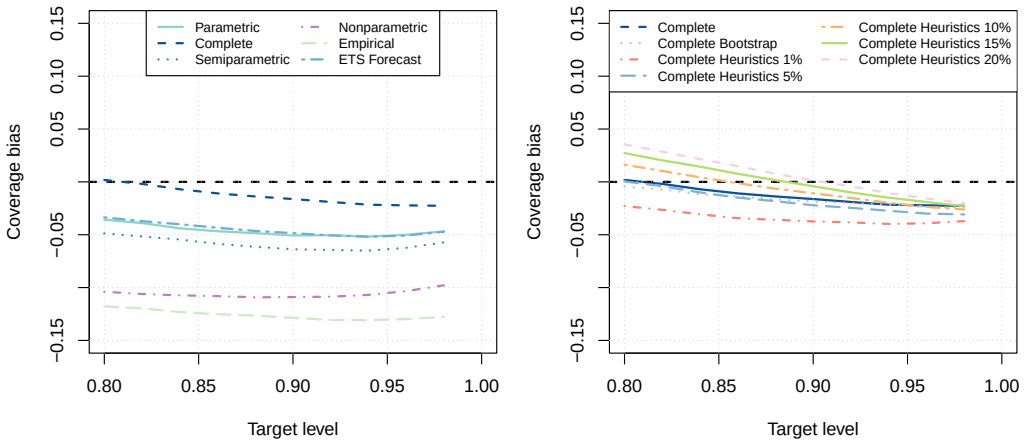
Figure 1: Coverage bias for all approaches under consideration

Figure 1 summarises performance of all approaches for different coverage levels from 2% to 98%. The ideal approach should lie on the black line, where the achieved level coincides with the target one. We can say that overall, none of the approaches achieved that, although on the lower levels (approximately until 60%), the parametric intervals are doing well. Interestingly enough, the nonparametric and empirical intervals tend to under perform, containing consistently fewer observations than needed, while heuristic approaches tend to over perform, especially in the middle, around 50%. One more thing to note is that the parametric intervals from `adam()` and `ets()` are performing very similarly, something we expected to see anyway.

In practice, the higher coverage (roughly more than 80%) is more important, so we focus on the right-hand side of the plot, which is shown in Figure 2. Figure 2a summarises performance of the conventional approaches together with the Hessian-based method (for reference), while Figure 2b shows how the proposed approaches perform.

We see that when it comes to higher target levels, all the conventional approaches tend to under perform and produce intervals that contain fewer than needed observations. The situation with the scenario-based method proposed in this paper is slightly better, especially on the levels lower than 90%, but deteriorates, when the higher coverage is needed. It also becomes apparent that the proportion used in the heuristics approach should vary depending on the required confidence level, e.g. the 10% one performs very well in case of 85% confidence level, while the 20% one does better on the 90% confidence level. The 1% heuristics is doing consistently poorly in this specific part of confidence level, not being able to get close to the target level.

To understand better the performance of approaches in terms of coverage, we look in more



(a) Conventional approaches (b) Scenario-based approaches

Figure 2: Coverage for 80% and higher confidence levels.

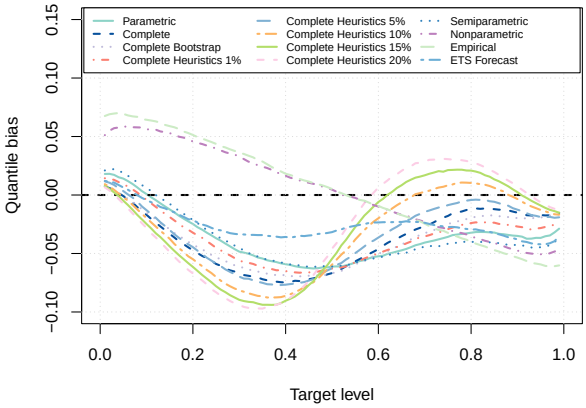
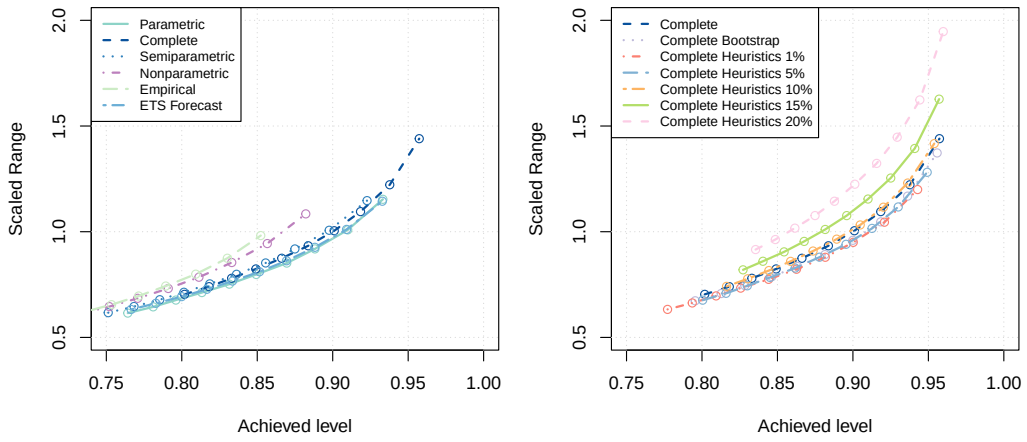


Figure 3: Quantile bias for all the approaches.

detail to the quantile performance of approaches, measuring the proportion of cases lying below each specific quantile from 1% to 99% and then subtracting the target value from each of them. The results are shown in Figure 3. There are several things to note based on the plot. First, none of the approaches produced quantiles that would be consistently close to the target for all confidence levels: some approaches do better for lower quantiles, the others do better for the upper ones. Second, nonparametric and empirical intervals perform very similar, having narrower intervals than expected, which is reflected in positive quantile bias for the lower levels and the negative for the higher ones. Third, the parametric intervals, the intervals from the `ets()` function and the semiparametric ones tend to underestimate the quantiles,

producing the values consistently lower than required (i.e. having negative quantile bias). Forth, when it comes to the "Complete" approaches, they tend to substantially underestimate the quantile for the levels below 50% and do slightly better, when the higher levels are required (e.g. above 80%), outperforming there the conventional approaches. The latter happens because the additional source of uncertainty is captured and thus upper quantiles lie higher than in case of the conventional approaches. Finally, the fact that the majority of approaches tend to underestimate the quantiles (producing negative quantile biases) means that in many cases the point forecasts are biased, because the intervals for the ETS models are typically constructed symmetrically around them. This could be due to the change of the tendencies in the holdout period and a lack of explanatory variables in the model. This cannot be fixed automatically and needs to be addressed on individual basis per model, per series.

Another visualisation that we find useful is presented in Figure 4. It shows the trade-off between coverage and range for the conventional approaches (Figure 4a) and the proposed ones (Figure 4b). The circles denote each of the confidence levels, starting from 80% and ending with 98%.



(a) Conventional approaches

(b) Scenario-based approaches

Figure 4: Coverage vs Range for 80% and higher confidence levels.

It becomes apparent (from Figure 4a) that the Nonparametric and Empirical intervals have higher range than the other conventional approaches, but achieve the lower levels than they do. The "complete" (Hessian-based) approach does better in terms of achieving higher levels at a low cost of increased range. When it comes to Figure 4b, the heuristic methods achieve higher

coverage at the high cost of the range of the intervals, e.g. the 20% heuristics is much closer to the nominal levels, but has the highest range, making it impractical to use in real life. The 10% heuristics and the Hessian-based method can be considered as the ones having a reasonable balance between the coverage and range: they tend to produce levels closer to the target ones than the methods with the lower values of heuristics proportions (as seen in Figure 2b), and do not have as high range as the levels with the higher heuristics proportions.

6. Conclusion

ETS, as any other dynamic model (such as ARIMA), tends to underestimate the uncertainty in the future, producing narrower prediction intervals than required. In this paper we argue that this happens due to the ignoring of one of the fundamental sources of uncertainty, namely the uncertainty coming from the estimates of parameters in sample.

We showed analytically what formulae can be used to produce more correct h-steps ahead variance conditional on all the available information in sample. Unfortunately, the usefulness of these formulae is limited to pure additive ETS only and to two separate cases: when the initials are estimated, but the smoothing parameters are known and vice versa.

We then developed a scenario-based approach, which relies on applying many ETS models with different parameters to the data to capture and propagate the uncertainty coming from their estimation to the final prediction intervals.

Finally, we conducted an experiment on M competition and Tourism competition data to see how the approach works in the automated fashion. We noticed that it failed in some cases of challenging time series, which we then dropped from the investigation. The results showed that the scenario-based approaches tend to produce wider intervals, which guarantees a higher coverage, but in some cases also means that the lower quantiles are miscalibrated. In fact, we found that none of the approaches for interval construction based on ETS produces a calibrated and unbiased quantile forecasts. We argue that this could be due to the missing information in the models (e.g. the lack of explanatory variables), but the fix of this issue lies outside of the scope of this paper.

For the future work, it would be interesting to investigate how the proposed approach works in conjunction with forecasts combinations, e.g. based on AIC weights, similar to how it was done by Kolassa (2011). Furthermore, the developed approach can be applied to any model and, for example, can be used for the generation of quantiles from ARIMA. Investigating how

it works in that case would be another possible direction for future research.

References

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

Bergmeir, C., R. J. Hyndman, and J. M. Benítez. Bagging exponential smoothing methods using STL decomposition and Box-Cox transformation. *International Journal of Forecasting* (2), 303–312.

Chatfield, C. (1996). Model uncertainty and forecast accuracy. *Journal of Forecasting 15*(7), 495–508.

Cleveland, R. B., W. S. Cleveland, J. E. McRae, and I. Terpenning (1990). STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics 6*(1), 3–73.

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

Hyndman, R. J. and Y. Khandakar (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software 26*(3), 1–22.

Hyndman, R. J., A. B. Koehler, J. K. Ord, and R. D. Snyder (2008). *Forecasting with Exponential Smoothing*. Springer Berlin Heidelberg.

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

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

Kolassa, S. (2022, oct). Commentary on the M5 forecasting competition. *International Journal of Forecasting 38*(4), 1562–1568.

Lee, Y. S. and S. Scholtes (2014). Empirical prediction intervals revisited. *International Journal of Forecasting 30*, 217–234.

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

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

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

Makridakis, S., E. Spiliotis, and V. Assimakopoulos (2022). The M5 competition: Background, organization, and implementation. *International Journal of Forecasting 38*(4).

Petropoulos, F., R. J. Hyndman, and C. Bergmeir (2018). Exploring the sources of uncertainty: Why does bagging for time series forecasting work? *European Journal of Operational Research 268*(2), 545–554.

Petropoulos, F. and N. Kourentzes (2015, jun). Forecast combinations for intermittent demand. *Journal of the Operational Research Society 66*(6), 914–924.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Svetunkov, I. (2023a). *Forecasting and Analytics with the Augmented Dynamic Adaptive Model (ADAM)*. Chapman and Hall/CRC.

Svetunkov, I. (2023b). *smooth: Forecasting Using State Space Models*. R package version 3.2.2.

Svetunkov, I., H. Chen, and J. E. Boylan (2023, feb). A new taxonomy for vector exponential smoothing and its application to seasonal time series. *European Journal of Operational Research 304*(3), 964–980.

Taylor, J. W. and D. W. Bunn (1999). A Quantile Regression Approach to Generating Prediction Intervals. *Management Science 45*(2), 225–237.

Weller, M. and S. Crone (2012). Supply Chain Forecasting: Best Practices and Benchmarking Study. Technical report, Lancaster Centre for Forecasting.