Bagging Exponential Smoothing Methods using STL Decomposition and Box-Cox Transformation

C. Bergmeir[1], Rob J. Hyndman[2], and J. M. Benítez[1]

[1]Departamento de Ciencias de la Computación e Inteligencia Artificial, ETSIIT, CITIC-UGR, Universidad de Granada

[2]Department of Econometrics & Business Statistics, Monash University, Clayton, Australia

Forecasting with exponential smoothing (ETS)

Why use exponential smoothing (ETS)?

- More than 50 years of widespread use
- Still one of the most practically relevant forecasting methods available
- Simple, and able to adapt to many situations
- It has a solid theoretical foundation in ETS state-space models
- Competitive results in the M3 forecasting competition
- The `forecast` package in the programming language R: a fully automatic software for fitting ETS models

⇒ Any attempt to improve ETS point-forecast accuracy seems a worthwhile endeavor

Bootstrap aggregating (bagging)

- Popular method in Machine Learning to improve the accuracy of predictors
- An ensemble of predictors is estimated on bootstrapped versions of the input data
- The output of the ensemble is calculated by averaging
- This often yields better point predictions, as results are more robust

Bagging for exponential smoothing

- Our input data are time series

$\Rightarrow$ We have to take into account

  - Serial dependence
  - Non-stationarity

$\Rightarrow$ We resolve these issues by applying:

  - Box-Cox transformation
  - STL/Loess decomposition
  - Moving block bootstrap (MBB) on the residuals of the decomposition
  - ETS on a set of reconstructed series
  - Averaging of the resulting point forecasts

- We evaluate the procedure on the M3 data, with competitive results

Holt-Winters method

The Holt-Winters method with additive trend and seasonality is defined by:

$$\begin{aligned}
\ell_t &= \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1} \\
s_t &= \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} \\
\hat{y}_{t+h|t} &= \ell_t + hb_t + s_{t-m+h_m^+}
\end{aligned}$$

- ℓ_t, b_t, and s_t are the level, trend, and seasonal index at time t.
- $\{y_t\}_{t=1,\dots,n}$ are the observed data.
- α, β^*, γ are the smoothing parameters, and m is the length of the seasonal cycle.
- $\hat{y}_{t+h|t}$ is a forecast at time t with horizon h, and $h_m^+ = \lfloor (h-1) \bmod m \rfloor + 1$.

Exponential smoothing model family

- Family of ETS models
- Models are distinguished by type of error (E), trend (T), and seasonality (S)
- Error can be additive or multiplicative
- Trend can be non-existent, additive, multiplicative, or damped additive/multiplicative
- Seasonality can be non-existent, additive, or multiplicative
- 30 models in total
- Function `ets` in R:
  - Smoothing parameters and initial conditions are optimized using maximum likelihood and a simplex optimizer
  - Function automatically selects a model based on AIC or AICc

Box-Cox Transformation

This is a popular transformation to stabilize the variance of a time series, defined as:

$$w_t = \begin{cases} \log(y_t), & \lambda = 0; \\ (y_t^\lambda - 1)/\lambda, & \lambda \neq 0. \end{cases}$$

- Depending on lambda, the transformation is:
    - the identity ($\lambda = 1$)
    - the logarithm ($\lambda = 0$)
    - a transformation in between
- We choose lambda using the procedure of Guerrero (1993).

Time series decomposition

For non-seasonal time series (yearly time series), we use the loess method:

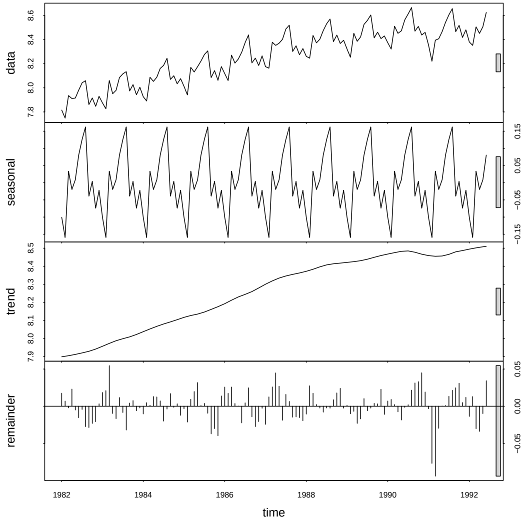
- A smoothing method based on local regressions
- Decomposes the time series into trend and remainder components
- A polynomial is fitted to a weighted neighborhood of each point
- We use $d = 1$ as degree of the polynomial
- Available in R as function `loess`

Time series decomposition (2)

For seasonal time series, we use STL decomposition:

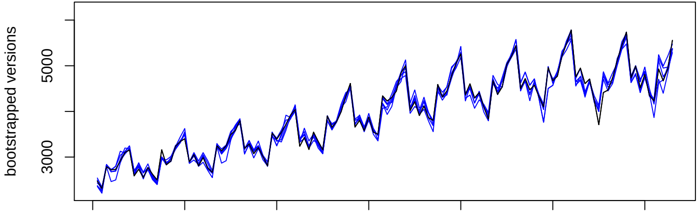
- Seasonal-trend decomposition based on loess
- The loess method is applied repeatedly in different ways
- Series is divided into trend, seasonal part, and remainder
- The division is additive
- Available in R in function `stl`

Time series decomposition (3)



Bootstrapping the remainder

- Time series are autocorrelated $\Rightarrow$ Moving Block Bootstrap (MBB) is used
- Stationarity is achieved by bootstrapping the remainder of stl/loess
- We use block sizes of $l = 8$ for yearly and quarterly data, and $l = 24$ for monthly data, to capture any remaining seasonality



Overall Procedure

```
procedure BOOTSTRAP(ts, num.boot)
    λ ← BoxCox.lambda(ts)    ▷ use automatic procedure to determine λ
    ts.bc ← BoxCox(ts, λ)
    if ts is seasonal then
        [trend, seasonal, remainder] ← stl(ts.bc)
    else
        seasonal ← 0
        [trend, remainder] ← loess(ts.bc)
    end if
    recon.series[1] ← ts    ▷ add the original series as a sample
    for i in 2 to num.boot do
        boot.sample[i] ← MBB(remainder)
        recon.series.bc[i] ← trend + seasonal + boot.sample[i]
        recon.series[i] ← InvBoxCox(recon.series.bc[i], λ)
    end for
    return recon.series
end procedure
```

Used methods

Benchmark methods:

ETS: The original ETS method choosing among models using AICc.

AdditiveETS.BC: Box-Cox transformation with the same value for λ as in the bootstrapping is applied before using ETS.
ETS chooses only from the additive models

Bagging procedures:

BaggedETS: For each of the bootstrapped series (we use 30 series), *ETS* is applied. Then, for the forecasts, the outcome of all methods is averaged using the mean.

BaggedETS.BC: Same as *BaggedETS*, but using *AdditiveETS.BC* as the base procedure.

Evaluation methodology

Data from the M3 forecasting competition:

- 645 yearly series, 756 quarterly series, and 1428 monthly series. In total, 2829 series are used.
- Forecast horizons: 6 values for yearly series, 8 values for quarterly series, 18 values for monthly series.
- Results of original M3 participants are used as benchmarks.

Error measures used:

$$\text{sMAPE} = \text{mean}\left(200\frac{|y_t - \hat{y}_t|}{|y_t| + |\hat{y}_t|}\right)$$

$$\text{MASE} = \frac{\text{mean}\left(|y_t - \hat{y}_t|\right)}{\text{mean}\left(|y_i - y_{i-m}|\right)}$$

Errors are averaged over all horizons, MASE uses seasonal naïve method as benchmark.

Evaluation methodology (2)

Statistical testing:

- Aligned Friedman rank-sum test for multiple comparisons
- Hochberg post-hoc procedure for comparisons against a control method
- In a first step, we compare the proposed methods to each other
- In a second step, we compare the best of the proposed methods to the M3 benchmarks
- We report results for sMAPE

Results on yearly data

	Rank sMAPE	Mean sMAPE	Rank MASE	Mean MASE
ForcX	12.022	16.480	12.007	2.769
AutoBox2	12.370	16.593	12.388	2.754
RBF	12.381	16.424	12.389	2.720
Flors.Pearc1	12.471	17.205	12.476	2.938
THETA	12.530	16.974	12.557	2.806
ForecastPro	12.615	17.271	12.636	3.026
ROBUST.Trend	12.769	17.033	12.809	2.625
PP.Autocast	12.800	17.128	12.788	3.016
DAMPEN	12.841	17.360	12.828	3.032
BaggedETS	12.853	17.492	12.840	2.973
COMB.S.H.D	12.960	17.072	12.910	2.876
SMARTFCS	13.310	17.706	13.333	2.996
AdditiveETS.BC	13.333	18.357	13.389	3.557
BaggedETS.BC	13.408	18.258	13.411	3.478
HOLT	13.554	20.021	13.565	3.182
WINTER	13.554	20.021	13.565	3.182
ETS	13.884	18.655	13.963	3.532
ARARMA	14.025	18.356	14.129	3.481
Flors.Pearc2	14.029	17.843	14.054	3.016
B.J.auto	14.068	17.726	14.041	3.165
Auto.ANN	14.420	18.565	14.394	3.058
AutoBox3	14.602	20.877	14.592	3.177
THETAsm	14.684	17.922	14.598	3.006
AutoBox1	14.931	21.588	14.930	3.679
NAIVE2	15.270	17.880	15.195	3.172
SINGLE	15.316	17.817	15.212	3.171

Results for the yearly series, ordered by the first column, which is the average rank of sMAPE. The other columns show mean sMAPE, average rank of MASE, and mean of MASE.

Results on yearly data (2)

Method	p_{Hoch}
BaggedETS	–
AdditiveETS.BC	0.023
BaggedETS.BC	0.023
ETS	0.002

Results of statistical testing for yearly data, using only exponential smoothing methods. Adjusted p-values calculated from the aligned Friedman test with Hochberg's post-hoc procedure are shown. A horizontal line separates the methods that perform significantly worse than the best method from the ones that do not.

Results on yearly data (3)

Method	p_{Hoch}
ForcX	–
AutoBox2	0.598
RBF	0.598
ROBUST.Trend	0.598
THETA	0.598
ForecastPro	0.598
Flors.Pearc1	0.598
PP.Autocast	0.574
DAMPEN	0.344
BaggedETS	0.344
COMB.S.H.D	0.085
SMARTFCS	0.056
ARARMA	0.001
B.J.auto	5.20×10^{-4}
WINTER	5.16×10^{-5}
HOLT	5.16×10^{-5}
Flors.Pearc2	5.16×10^{-5}
Auto.ANN	1.37×10^{-5}
AutoBox3	2.81×10^{-7}
SINGLE	7.79×10^{-8}
THETAsm	6.94×10^{-8}
NAIVE2	5.69×10^{-8}
AutoBox1	2.62×10^{-11}

Results of statistical testing for yearly data, using BaggedETS and the original results of the M3. Adjusted *p*-values calculated from the aligned Friedman test with Hochberg's post-hoc procedure are shown. A horizontal line separates the methods that perform significantly worse than the best method from the ones that do not.

Results on quarterly data

	Rank sMAPE	Mean sMAPE	Rank MASE	Mean MASE
THETA	11.493	8.956	11.483	1.087
COMB.S.H.D	12.196	9.216	12.187	1.105
ROBUST.Trend	12.489	9.789	12.500	1.152
DAMPEN	12.692	9.361	12.697	1.126
ForcX	12.843	9.537	12.847	1.155
PP.Autocast	12.847	9.395	12.866	1.128
B.J.auto	13.161	10.260	13.169	1.188
ForecastPro	13.192	9.815	13.208	1.204
HOLT	13.236	10.938	13.171	1.225
RBF	13.267	9.565	13.251	1.173
WINTER	13.413	10.840	13.355	1.217
BaggedETS	13.425	10.030	13.438	1.192
ARARMA	13.426	10.186	13.393	1.185
BaggedETS.BC	13.493	10.059	13.523	1.209
Flors.Pearc1	13.500	9.954	13.511	1.184
AutoBox2	13.513	10.004	13.552	1.185
AdditiveETS.BC	13.671	9.895	13.704	1.224
ETS	13.713	9.838	13.697	1.214
Auto.ANN	13.961	10.199	13.976	1.241
THETAsm	14.238	9.821	14.217	1.211
SMARTFCS	14.249	10.153	14.296	1.226
Flors.Pearc2	14.335	10.431	14.415	1.255
AutoBox3	14.458	11.192	14.386	1.272
SINGLE	14.637	9.717	14.610	1.229
AutoBox1	14.729	10.961	14.735	1.331
NAIVE2	14.823	9.951	14.814	1.238

Results for the quarterly series, ordered by the first column, which is the average rank of sMAPE.

Results on quarterly data (2)

Method	p_{Hoch}
BaggedETS.BC	–
BaggedETS	0.777
ETS	0.777
AdditiveETS.BC	0.777

Results of statistical testing for quarterly data, using only exponential smoothing methods. None of the results is statistically significant (so no horizontal line is drawn).

Results on quarterly data (3)

Method	p_{Hoch}
THETA	–
COMB.S.H.D	0.148
DAMPEN	0.031
PP.Autocast	0.031
ROBUST.Trend	0.031
ForcX	0.018
RBF	0.002
ForecastPro	7.66×10^{-4}
ARARMA	1.19×10^{-4}
Flors.Pearc1	1.19×10^{-4}
B.J.auto	1.07×10^{-4}
AutoBox2	1.07×10^{-4}
BaggedETS	1.60×10^{-5}
HOLT	1.56×10^{-5}
WINTER	6.46×10^{-6}
THETAsm	3.54×10^{-6}
SINGLE	1.63×10^{-7}
Auto.ANN	1.16×10^{-7}
NAIVE2	9.64×10^{-9}
SMARTFCS	8.33×10^{-9}
Flors.Pearc2	3.12×10^{-10}
AutoBox3	5.43×10^{-11}
AutoBox1	1.91×10^{-11}

Results of statistical testing for quarterly data, using BaggedETS.BC and the original results of the M3. A horizontal line separates the methods that perform significantly worse than the best method from the ones that do not. We see that only the COMB.S.H.D is not worse with statistical significance than the THETA method.

Results on monthly data

	Rank sMAPE	Mean sMAPE	Rank MASE	Mean MASE
BaggedETS.BC	11.178	13.739	11.148	0.852
THETA	11.474	13.892	11.405	0.858
ForecastPro	11.578	13.898	11.597	0.848
BaggedETS	12.047	14.347	11.925	0.854
AdditiveETS.BC	12.424	14.224	12.373	0.891
COMB.S.H.D	12.527	14.466	12.584	0.896
ETS	12.648	14.375	12.718	0.895
HOLT	12.834	15.795	12.794	0.909
ForcX	12.844	14.466	12.894	0.894
WINTER	13.117	15.926	13.100	1.165
RBF	13.302	14.760	13.323	0.910
DAMPEN	13.578	14.576	13.631	0.908
AutoBox2	13.677	15.731	13.738	1.082
B.J.auto	13.746	14.796	13.744	0.914
AutoBox1	13.790	15.811	13.800	0.924
Flors.Pearc2	13.915	15.186	13.947	0.950
SMARTFCS	13.933	15.007	13.833	0.919
Auto.ANN	13.975	15.031	13.992	0.928
ARARMA	14.155	15.826	14.164	0.907
PP.Autocast	14.256	15.328	14.340	0.994
AutoBox3	14.339	16.590	14.242	0.962
Flors.Pearc1	14.627	15.986	14.636	1.008
THETAsm	14.674	15.380	14.687	0.950
ROBUST.Trend	14.874	18.931	14.781	1.039
SINGLE	15.329	15.300	15.412	0.974
NAIVE2	16.157	16.891	16.190	1.037

Results for the monthly series, ordered by the first column, which is the average rank of sMAPE.

Results on monthly data (2)

Method	p_{Hoch}
BaggedETS.BC	–
BaggedETS	3.12×10^{-7}
AdditiveETS.BC	8.09×10^{-12}
ETS	1.04×10^{-17}

Results of statistical testing for monthly data, using only exponential smoothing methods. The BaggedETS.BC method clearly outperforms all other methods.

Results on monthly data (3)

Method	p_{Hoch}
BaggedETS.BC	–
THETA	0.087
ForecastPro	0.087
COMB.S.H.D	2.27×10^{-7}
ForcX	1.59×10^{-8}
HOLT	1.69×10^{-10}
RBF	1.36×10^{-10}
DAMPEN	9.02×10^{-12}
Auto.ANN	6.64×10^{-14}
WINTER	5.13×10^{-14}
B.J.auto	1.09×10^{-14}
Flors.Pearc2	6.43×10^{-15}
SMARTFCS	4.01×10^{-16}
AutoBox1	1.61×10^{-17}
AutoBox2	5.00×10^{-18}
PP.Autocast	2.12×10^{-18}
AutoBox3	3.46×10^{-24}
ARARMA	2.30×10^{-24}
THETAsm	6.77×10^{-25}
Flors.Pearc1	1.95×10^{-25}
SINGLE	9.72×10^{-35}
ROBUST.Trend	4.50×10^{-49}
NAIVE2	4.17×10^{-62}

Results of statistical testing for monthly data, using BaggedETS.BC and the original results of the M3. BaggedETS.BC performs best, and only the THETA and ForecastPro methods do not perform significantly worse. However, their p-values are close to 0.05 so that they are nearly significant.

Conclusions

- A method of bagging for exponential smoothing, using Box-Cox transformation, STL decomposition, and the moving block bootstrap has been presented.
- The bagged exponential smoothing methods consistently outperform the basic exponential smoothing methods.
- On the monthly data of the M3 competition, the bagged exponential smoothing method with Box-Cox transformation and additive-only models is able to outperform, with statistical significance, all methods that took part in the competition.
- So, especially for monthly data this method can be recommended to be used in practice.
- Working paper: http://robjhyndman.com/working-papers/bagging-ets/

Thank you

Christoph Bergmeir

c.bergmeir@decsai.ugr.es