Forecasting and Analytics with ADAM

Ivan Svetunkov

TAFS

14th September 2022

Marketing Analytics and Forecasting | Lancaster University Management School

What is forecasting?

What is forecasting?

Forecasting is the process for producing forecasts (duh!)

Forecast is a scientifically justified assertion about possible states of an object in future.

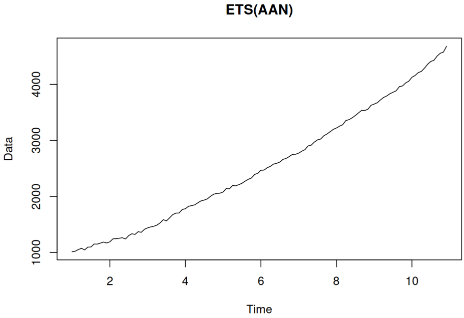
It's not the same as "foretelling" or "guessing".

The word "prediction" is typically used as a synonym (should it be?).

What is forecasting?

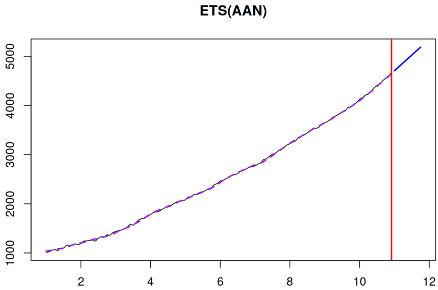
In practice, forecasting comes to finding structure in data and extrapolating it into the future.

What will happen in the future?



What is forecasting?

Probably something like this:



Classical decomposition

So, how do we forecast?

One of the simpler approaches – decompose time series y_t into several components:

1. Level l_t,
2. Trend (growth) b_t,
3. Seasonality s_t,
4. Error ϵ_t

Each time series can have a combination of the four in different form.

Classical decomposition

We can add components or multiply them.

Pure additive model:

$$y_t = l_{t-1} + b_{t-1} + s_{t-m} + \epsilon_t$$

Pure multiplicative model:

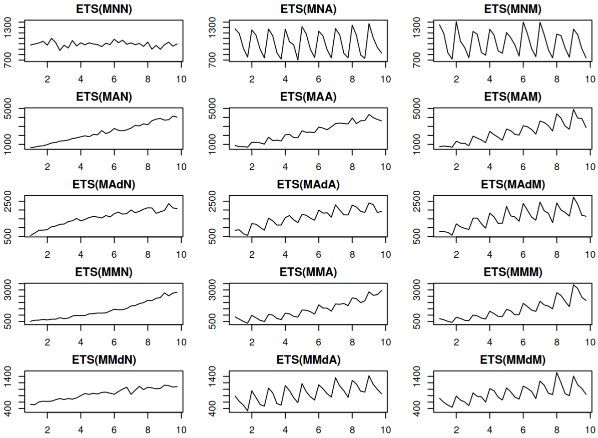
$$y_t = l_{t-1} b_{t-1} s_{t-m} (1 + \epsilon_t)$$

A mixed model:

$$y_t = (l_{t-1} + b_{t-1}) s_{t-m} (1 + \epsilon_t)$$

ETS

This is the basis for the ETS taxonomy (Error-Trend-Seasonality by Hyndman et al., 2008).



ARIMA, Autoregression

An alternative to capture the structure is to find relations over time.

Do sales today depend on sales in the past?

No. But let's assume that they do.

We can then say, for example that on average 90% of yesterday sales will happen today:

$y_t = 0.9 y_{t-1} + \epsilon_t$

This is an AR(1) model.

ARIMA, Autoregression

We can have as many past values as we want in AR(p):

$$y_t = \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \epsilon_t,$$

where ϕ_j is the j-th parameter of the model.

ARIMA, Moving Average

We can also have a structure, which correlates with past noise.

$$y_t = 0.3\epsilon_{t-1} + \epsilon_t$$

(white noise leads to 30% increase in sales?!)

This is an MA(1).

And we can have q past elements to get MA(q):

$$y_t = \theta_1\epsilon_{t-1} + ... + \theta_q\epsilon_{t-q} + \epsilon_t,$$

where θ_j is the j-th parameter of the model.

ARIMA, Integration (differences)

Some structure maybe captured if we switch to differences.

e.g. we predict changes in sales:

$$y_t - y_{t-1} = \epsilon_t$$

This is I(1).

We can also take differences of differences to get I(2).

Uniting AR(p), I(d) and MA(q), we get ARIMA(p,d,q) (Box and Jenkins, 1976).

Regression

We can also have some external information.

e.g. sales are driven by price changes, promotional activities, income changes.

We can capture the structure via the regression:

$$y_t = a_0 + a_1 x_{1,t} + \ldots + a_k x_{k,t} + \epsilon_t$$

where a_j is the j-th parameter for the $x_{j,t}$ explanatory variable.

ETS, ARIMA and Regression

There are other approaches.

But these are the three fundamental ones.

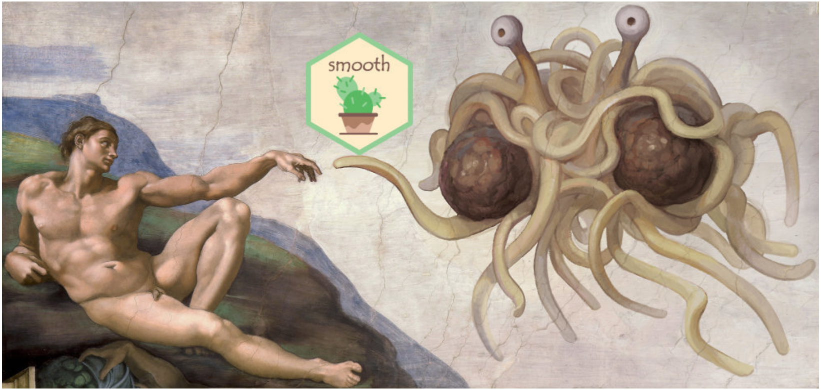
Almost all textbooks tell you that ETS, ARIMA and Regression are distinct models.

They are formulated differently.

In some cases some of them can be combined in one model: ETSX / ARIMAX.

But in general they are not comparable (e.g. via information criteria).

Introducing ADAM



What is ADAM?

ADAM is Augmented Dynamic Adaptive Model.

It is a Single Source of Error state space model, implementing:

1. Exponential Smoothing (ETS);
2. ARIMA;
3. Regression and TVP regression;
4. Combination of (1), (2) and (3);
5. Components, variables and orders selection;
6. Normal and non-normal distributions;
7. Advanced and custom losses;
8. ...

What is ADAM?

How do we achieve this?

Single Source of Error state space model (modified version of Hyndman et al., 2008):

$$\begin{aligned} y_t &= w(\mathbf{v}_{t-\mathbf{l}}) + r(\mathbf{v}_{t-\mathbf{l}})\epsilon_t \\ \mathbf{v}_t &= f(\mathbf{v}_{t-\mathbf{l}}) + g(\mathbf{v}_{t-\mathbf{l}})\epsilon_t \end{aligned} . \tag{1}$$

This model underlies all ETS and ARIMA models and allows incorporating explanatory variables.

What is ADAM?

ADAM is the instrument for:

- Benchmarking;
- Experimenting;
- Prototyping;
- Research.

All the details are summarised in the online textbook:
`https://openforecast.org/adam/`

What is ADAM?

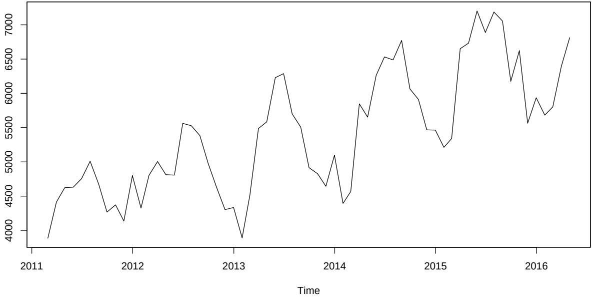
Instead of going through all the features, we will consider several case studies:

1. Fast Moving Consumer Goods (FMCG);
2. Promotional modelling;
3. Intermittent demand.

Fast Moving Consumer Goods

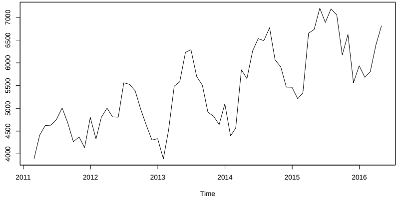
Fast Moving Consumer Goods

Sales of household product on a store level.



Use last year as the test set.

Fast Moving Consumer Goods



Things to note:

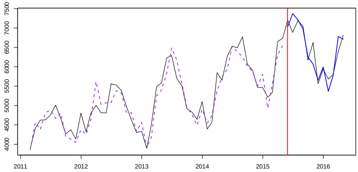
- The level changes over time,
- We have seasonality,
- Both seem to evolve over time

Use ETS Hyndman et al. (2008).

Fast Moving Consumer Goods

Use ETS with automatic components selection:

```
adamModel <- adam(y, h=12, holdout=TRUE, distribution="dnorm")
```



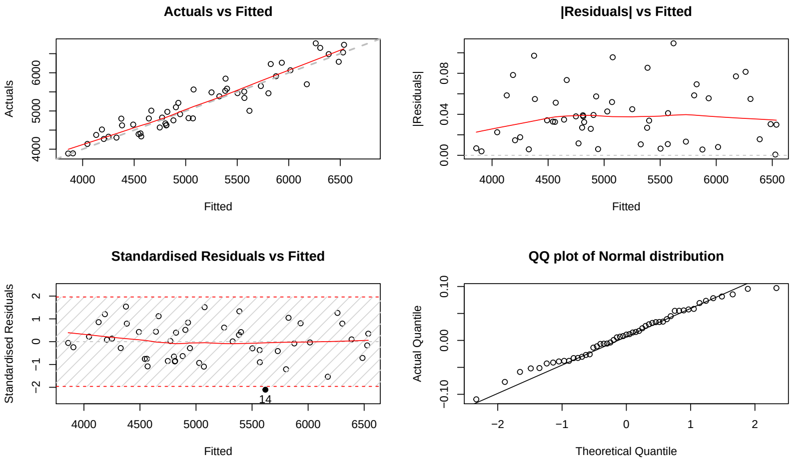
ETS(M,N,M):

- $\alpha = 0.882$, $\gamma = 0.003$;
- AICc=761.453
- For the holdout: MASE=0.790, RMSSE=0.739.

What about the analysis of the residuals?

```
plot(adamModel)
```

Fast Moving Consumer Goods



Fast Moving Consumer Goods

Observation number 14 is beyond the bounds.

This is fine, because 5% of values should lie outside.

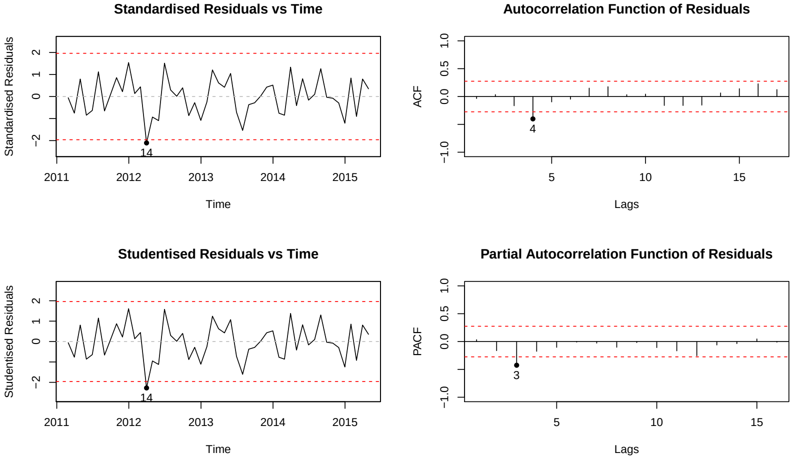
LOESS line on fitted vs Residuals is not straight.

But this could be because of randomness.

Further analysis:

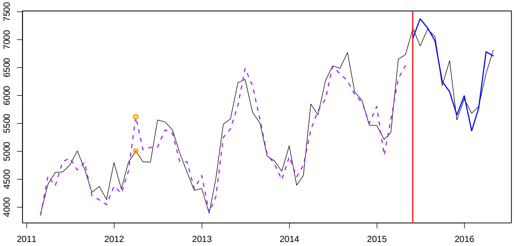
```
plot(adamModel, c(8,9,10,11))
```

Fast Moving Consumer Goods



Fast Moving Consumer Goods

- Observation 14 seems not random, it is followed by several other;
- Lags 3 and 4 PACF / ACF are significant, but this could be due to randomness;



Fast Moving Consumer Goods

Just to check, create the dummy variable for the outlier:

```
xreg <- cbind(y=y,x=0)
xreg$x[14] <- 1
adamModel2 <- adam(xreg, lags=12, h=12, holdout=T,
distribution="dnorm")
```

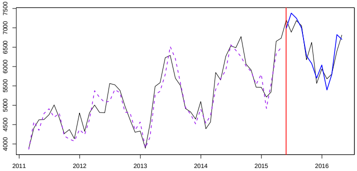
Fast Moving Consumer Goods

Just to check, create the dummy variable for the outlier:

```
xreg <- cbind(y=y,x=0)
xreg$x[14] <- 1
adamModel2 <- adam(xreg, lags=12, h=12, holdout=T,
distribution="dnorm")
```

ETSX(M,N,M):

- $\alpha = 0.908$, $\gamma = 0.004$;
- AICc=768.888
- For the holdout: MASE=0.785, RMSSE=0.750.

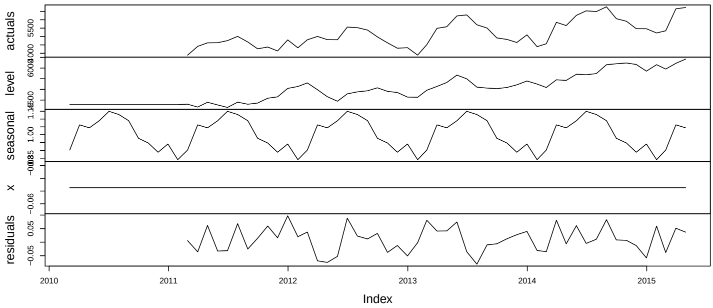


Not worth it! ETS(M,N,M) is good enough.

Fast Moving Consumer Goods

Time series decomposition according to ETS(M,N,M):

```
plot(adamModel, 12)
```



Fast Moving Consumer Goods, summary of the model

```
Model estimated using adam() function: ETS(MNM)
Response variable: data
Distribution used in the estimation: Normal
Loss function type: likelihood; Loss function value: 359.06
Coefficients:
             Estimate Std. Error Lower 2.5% Upper 97.5%
alpha          0.8816     0.1262     0.6259      1.0000
gamma          0.0031     0.1156     0.0000      0.1184
level       4282.1276   228.6296  3818.8800   4738.4743
seasonal_1     0.9025     0.0180     0.8661      0.9511
seasonal_2     1.0580     0.0194     1.0217      1.1067
seasonal_3     1.0445     0.0210     1.0082      1.0932
seasonal_4     1.0980     0.0223     1.0616      1.1466
seasonal_5     1.1486     0.0237     1.1122      1.1972
seasonal_6     1.1244     0.0244     1.0880      1.1730
seasonal_7     1.0916     0.0237     1.0552      1.1402
seasonal_8     0.9787     0.0210     0.9423      1.0273
seasonal_9     0.9490     0.0205     0.9126      0.9976
seasonal_10    0.8852     0.0182     0.8488      0.9338
seasonal_11    0.9394     0.0202     0.9030      0.9880

Sample size: 52
Number of estimated parameters: 15
Number of degrees of freedom: 37
Information criteria:
     AIC     AICc      BIC     BICc
748.1199 761.4533 777.3886 803.7302
```

Fast Moving Consumer Goods

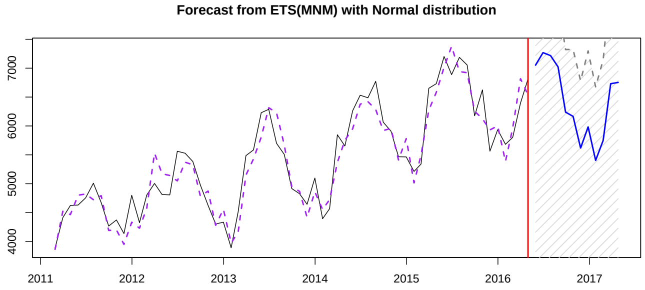
Final forecast (upper bound only):

```
plot(reforecast(adamModel, h=12, interval="prediction",
side="upper"))
```

Fast Moving Consumer Goods

Final forecast (upper bound only):

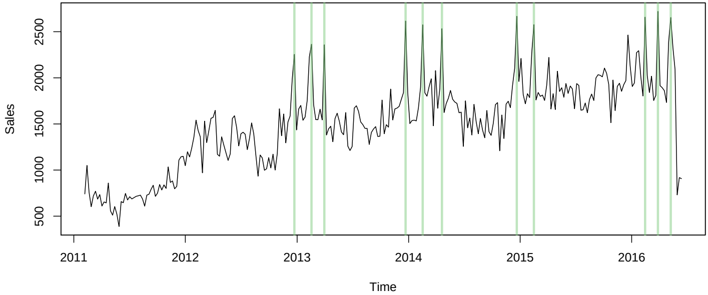
```
plot(reforecast(adamModel, h=12, interval="prediction",
side="upper"))
```



Promotional modelling

Promotional modelling

Weekly sales of a food product with some promotions

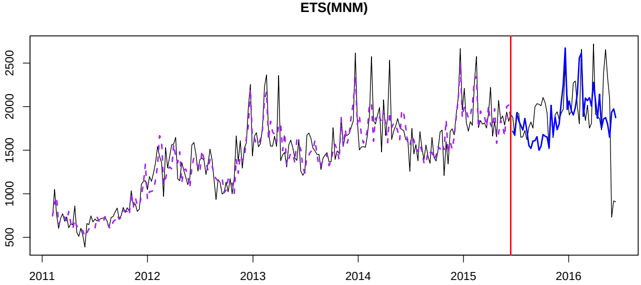


We'll use ETSX (similar to Koehler et al., 2012)

Promotional modelling

We will use automatic selection from pure models:

```
adamModel <- adam(y, "PPP", lags=c(1,52), h=52, holdout=TRUE)
```

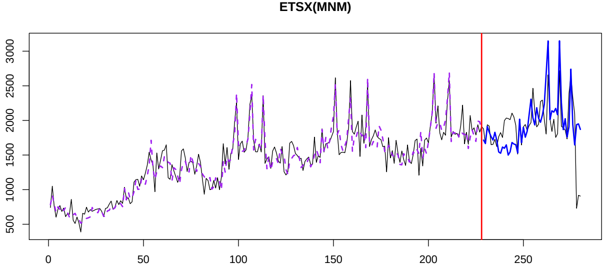


AICc=3072.125, MASE=1.508, RMSSE=1.453, Time=2.6 sec.

Promotional modelling

ETSX(M,N,M) with dummy for promotions and lags:

```
xreg <- data.frame(y, xregExpander(x, lags=-c(1,2), gaps="zero"))
adamModel <- adam(xreg, "MNM", lags=c(1,52), h=52, holdout=TRUE)
```

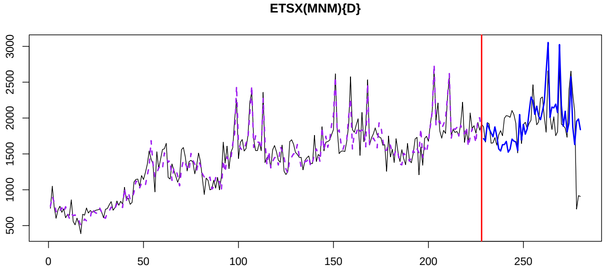


AICc=3040.972, MASE=1.526, RMSSE=1.446, Time=0.59 sec.

Promotional modelling

ETSX(M,N,M) with dynamic parameters:

```
adamModel <- adam(xreg, "MNM", lags=c(1,52), h=52, holdout=TRUE,
regressors="adapt")
```



AICc=3052.961, MASE=1.469, RMSSE=1.404, Time=0.59 sec.

Promotional modelling

Given AICc values, we should stick with the ETSX(M,N,M) with static regressors.

Dynamic regressors introduce additional uncertainty comming from smoothing parameters.

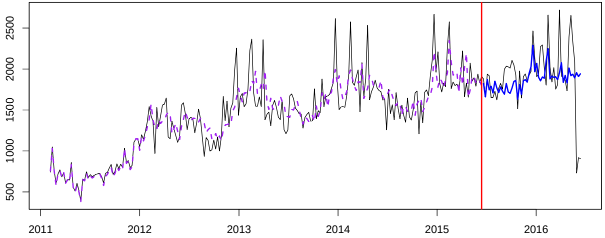
We could also use ARIMA and compare it with ETS:

```
adamModel <- adam(xreg, "NNN", lags=c(1,52),
orders=list(ar=c(3,2),i=c(2,1),ma=c(3,2),select=TRUE),
h=52, holdout=TRUE)
```

ARIMA is still work in progress!

Promotional modelling

SARIMA(0,1,3)(2,0,0)$_{52}$:

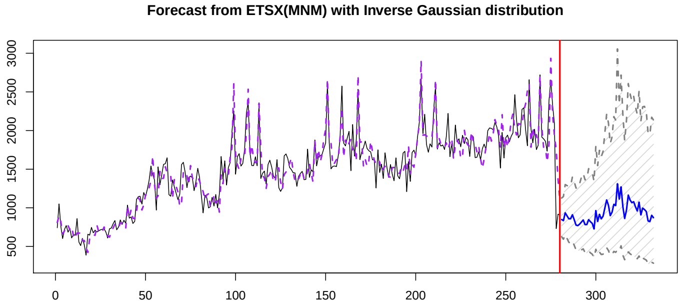


AICc=3493.406, MASE=1.242, RMSSE=1.312, Time=4.49 sec.

Worse than ETSX(M,N,M) based on AICc.

Promotional modelling

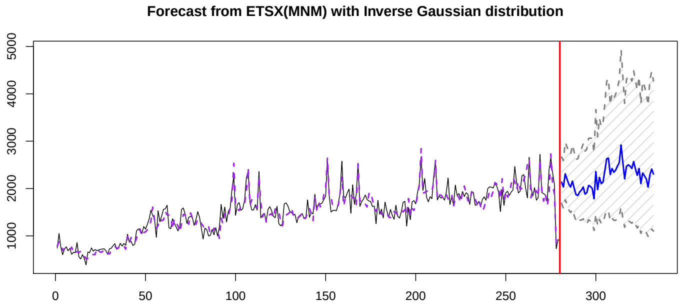
Final forecast from ETSX(M,N,M):



If the last three observations happened by chance, we could include dummies for them as well.

Promotional modelling

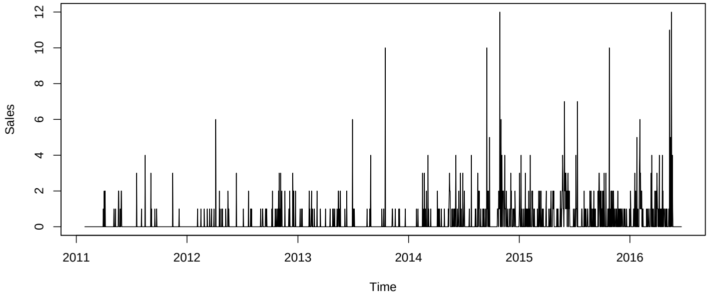
Alternative forecast from ETSX(M,N,M):



Intermittent demand

Intermittent demand

A “hobbies” product.



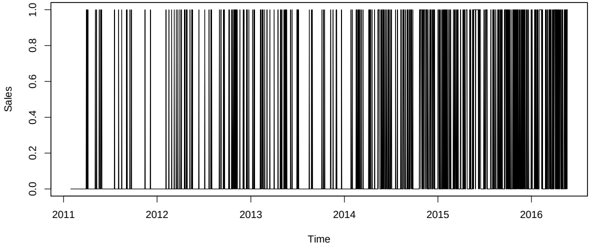
The iETS model is explained in **?**

Intermittent demand

This data can be split into two parts:

1. Demand occurrence part (0 / 1);
2. Demand sizes part;

Here how occurrence changes over time:



Intermittent demand

We see that the occurrence part evolves over time,

The probability of occurrence increases.

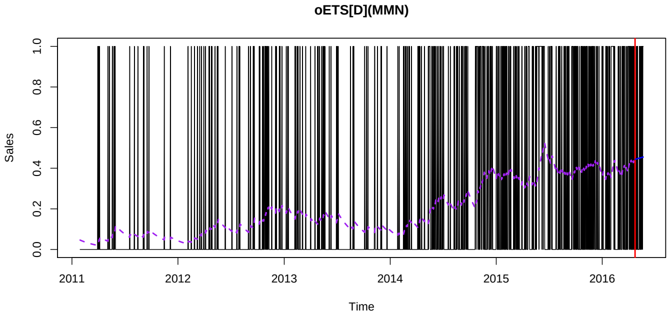
We can assume that it will increase in the holdout.

So we can use the model with trend to predict probability.

Intermittent demand

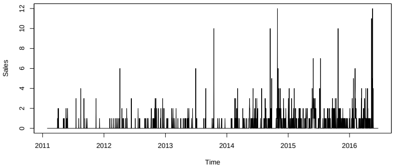
Use oETS(M,M,N) model for that:

```
oesModel <- oes(as.vector(y), "MMN", occurrence="direct",
h=28, holdout=TRUE)
plot(oesModel)
```



Intermittent demand

Now, what about the data itself?



The sizes increase over time. So trend would be suitable.

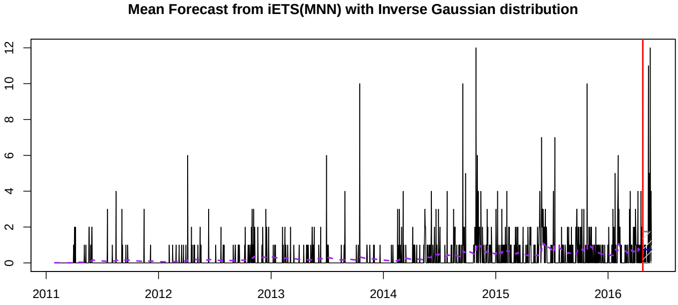
But let's select components automatically.

```
adamModel <- adam(y, "YYY", occurrence=oesModel,
h=28, holdout=TRUE)
```

Intermittent demand

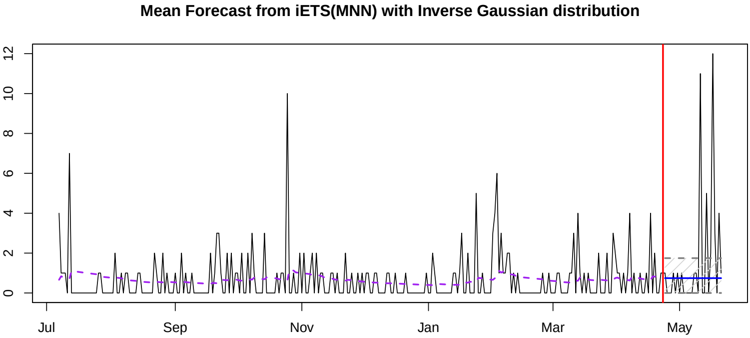
The final forecast (working and safety stocks):

```
plot(forecast(adamModel, h=28, interval="prediction",
nsim=10000, cumulative=TRUE, side="upper"))
```



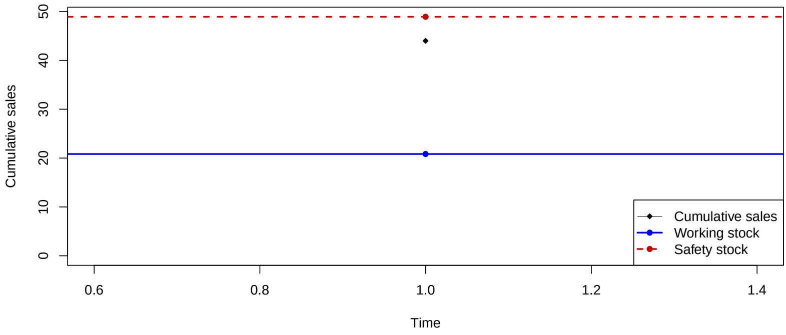
Intermittent demand

The final forecast zoomed in (mean over time values):



Intermittent demand

Still not very informative, so just compare cumulative sales over the 28 days with the forecasts



Conclusions

Conclusions

- ADAM is a new flexible model that supports many features;
- It was developed for demand forecasting, but can be used in other areas as well;
- It can be used for the standard problems instead of `es()` or `ets()`;
- It includes ARIMA in it;
- It can handle regressors;
- It can handle intermittent demand;
- It can handle multiple seasonal data;
- It can do a lot of other things.

Thank you for your attention!

Ivan Svetunkov

i.svetunkov@lancaster.ac.uk

https://forecasting.svetunkov.ru

https://openforecast.org/adam

twitter: @iSvetunkov

Marketing Analytics and Forecasting

Lancaster University Management School

References I

Box, G., Jenkins, G., 1976. Time series analysis: forecasting and control. Holden-day, Oakland, California.

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

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.

Svetunkov, I., Boylan, J., 2019. Multiplicative state-space models for intermittent time series.