One for all: forecasting intermittent and non-intermittent demand using one model

Ivan Svetunkov and John Boylan

Presentation at University of Bath

30th March 2017

Lancaster University Management School
Lancaster Centre for Forecasting

Introduction

Typical forecasting task in supply chain is to produce forecasts for many products.

Demand on each of the products may have its own characteristics and in general can be:

- non-intermittent;
- intermittent.

Introduction

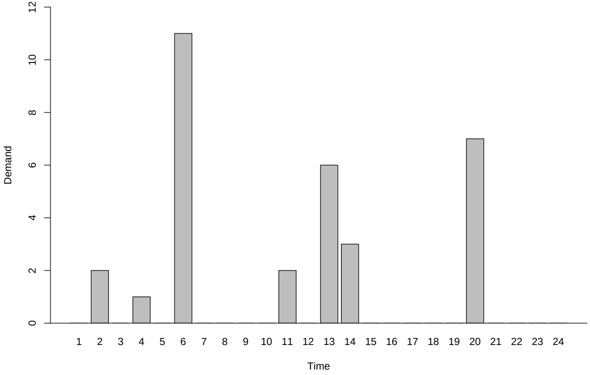


Figure: Example of intermittent data.

Introduction

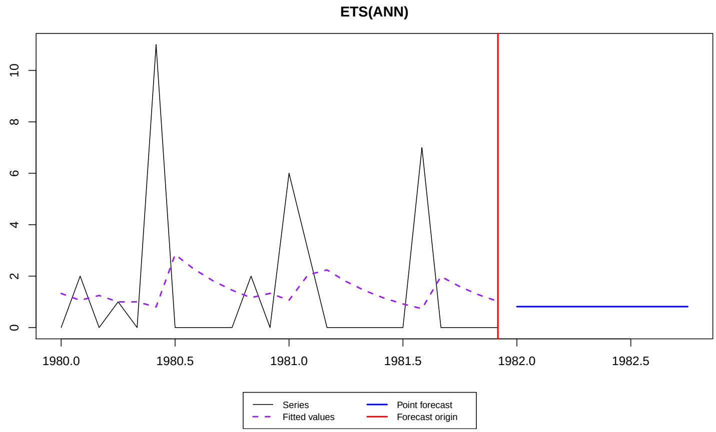


Figure: Simple Exponential Smoothing applied to the intermittent data.

Introduction

Intermittent data is considered as a separate case.

It is identified and then forecasted, usually using Croston (1972):

$$\begin{aligned} \hat{y}_t &= \frac{1}{\hat{q}_t}\hat{z}_t \\ \hat{z}_t &= \alpha_z z_{t-1} + (1-\alpha_z)\hat{z}_{t-1}, \\ \hat{q}_t &= \alpha_q q_{t-1} + (1-\alpha_q)\hat{q}_{t-1} \end{aligned} \quad (1)$$

where z_t are the demand sizes, q_t are the demand intervals,

α_z and α_q are the smoothing parameters.

Introduction

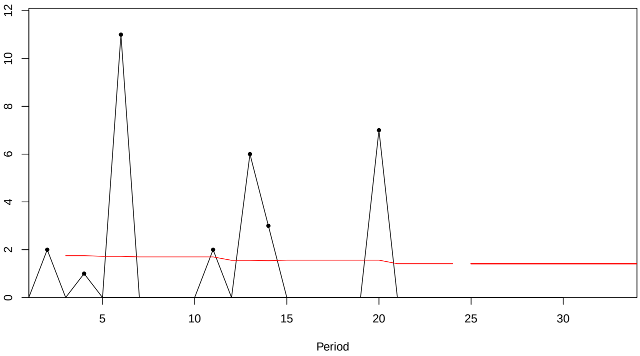


Figure: Intermittent data and Croston's forecast.

Introduction

We also have SBA (Syntetos and Boylan, 2005), TSB (Teunter et al., 2011), HES (Prestwich et al., 2014), INARMA etc.

All of them are separated from ETS / ARIMA / regression / etc.

Introduction

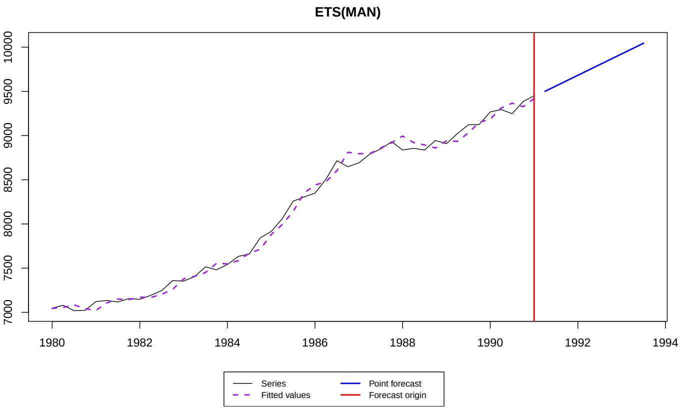


Figure: Non-intermittent data and a forecast.

Introduction

How to categorise the data?

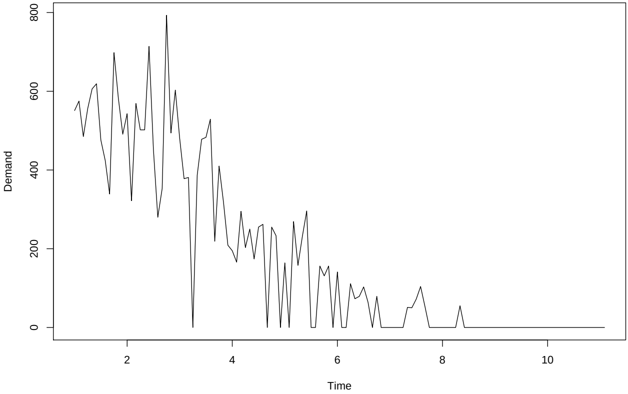
Johnston and Boylan (1996), Syntetos et al. (2005), Petropoulos and Kourentzes (2015)

BUT!

Products can change their characteristics...

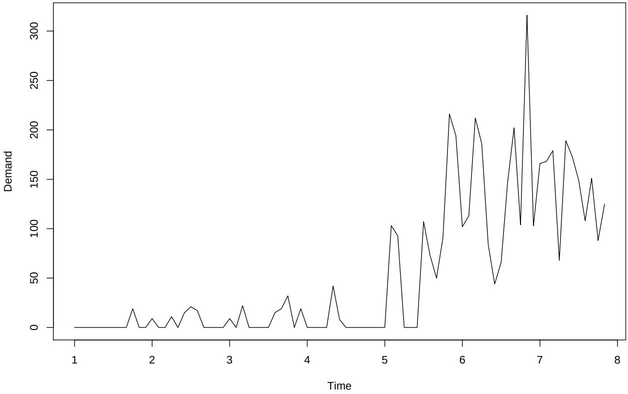
Introduction

Demand on a fast moving product may become obsolete...



Introduction

...or demand is just building up.



Problems

- Products can change their characteristics;
- Croston / TSB are based on SES.

Overall:

1. We need a model that could switch between intermittent / non-intermittent regimes;
2. We may need trend and / or seasonality;
3. We need to apply that model to a wide variety of data.

Intermittent state-space model (iSS)

Intermittent state-space model

The model is based on the original idea of Croston (1972):

$$y_t = o_t z_t, \tag{2}$$

where $o_t \sim \text{Bernoulli}(p_t)$ and z_t is a statistical model of our choice.

z_t can be ETS, ARIMA, regression, diffusion model, etc.

$o_t = 1$ means that there is a sale. $o_t = 0$ means no sale today.

If $o_t = 1$, for all t, then this is non-intermittent model.

General state-space (based on Hyndman et al. (2008))

General state-space model for iETS:

$$\begin{matrix} y_t = o_t \left(w(\mathbf{v}_{t-1}) + r(\mathbf{v}_{t-1}, \epsilon_t) \right) \\ v_t = f(\mathbf{v}_{t-1}) + g(\mathbf{v}_{t-1}, \epsilon_t) \end{matrix} . \tag{3}$$

$\mathbf{v}_t$ is the vector of states, w is the measurement function,

f is the transition function, g is the persistence function,

where $o_t \sim$ Bernoulli(p_t) and ϵ_t is the error term.

Intermittent state-space model

Multiplicative model is preferred (paper submitted to IJF).

Example. iETS(M,N,N) with time varying probability:

$$\begin{aligned} y_t &= o_t z_t \\ z_t &= l_{t-1}(1+\epsilon_t) \\ l_t &= l_{t-1}(1+\alpha\epsilon_t) \end{aligned} , \qquad (4)$$

$1+\epsilon_t \sim \text{logN}(0, \sigma^2)$, which means that z_t is always positive.

States are updated on every observation (potential demand).

But sales happen only when $o_t = 1$.

Underlies SES, when $o_t = 1$.

How to model the probability?

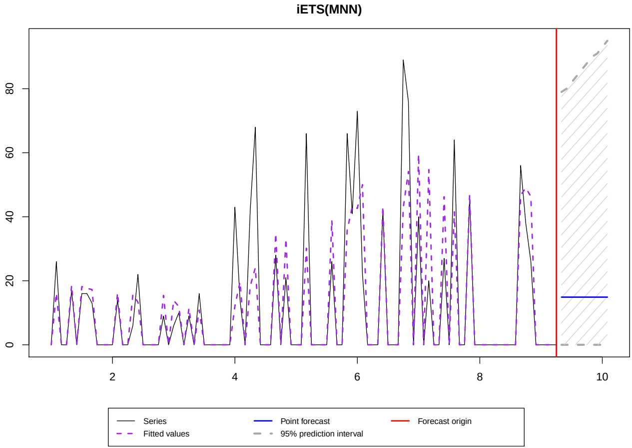
p_t has a statistical model of its own.

So far we have developed three models for p_t:

- Fixed probability model;
  $p_t = p$ for all t.
- Croston's model;
  $p_t = \frac{1}{1+q_t}$, where q_t is ETS(M,N,N).
- TSB model.
  $p_t \sim \text{Beta}(a_t, b_t)$, where a_t and b_t are ETS(M,N,N).

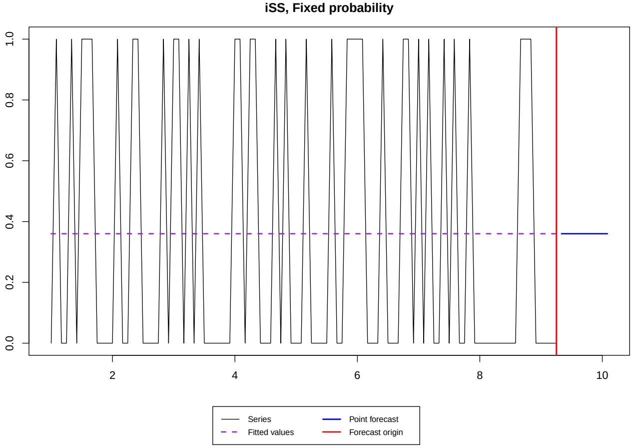
Examples

iETS(M,N,N) with fixed probability...



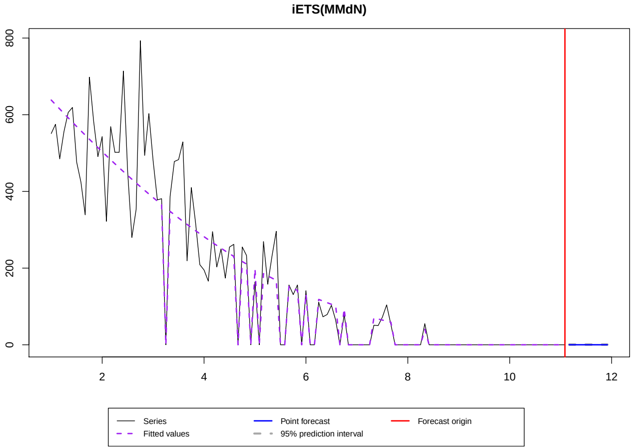
Examples

Fixed probability.



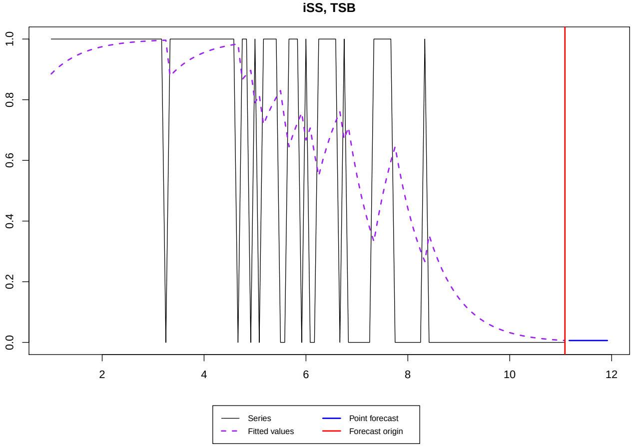
Examples

iETS(M,Md,N) with TSB and demand becoming obsolete



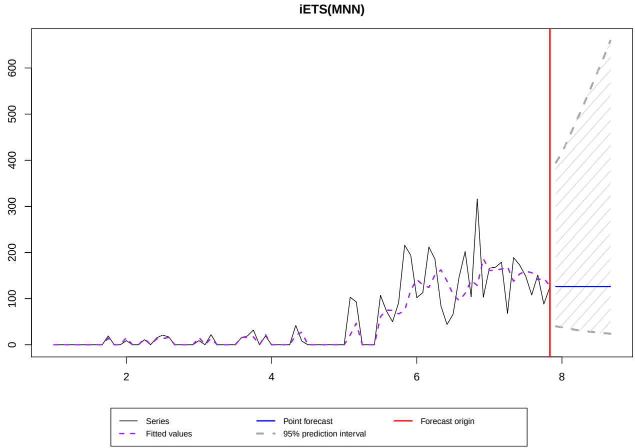
Examples

Time varying probability, TSB style.



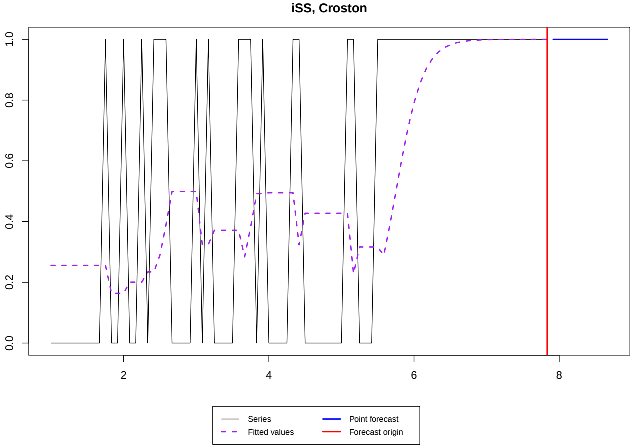
Examples

iETS(M,N,N) with Croston and building up level of demand...



Examples

Time varying probability, Croston's style.



Model selection in iSS

Model selection

Selection can be done in several directions:

1. z_t – select the best ETS model (error / trend / seasonality);
2. p_t – select the best model between Fixed / Croston / TSB;
3. p_t – select the best ETS model for Croston / TSB.

Here we discuss only (1) and (2).

Model selection

Concentrated log-ikelihood function for iETS model:

$$\begin{aligned} \ell(\theta, \hat{\sigma}_z^2 | \mathbf{Y}) = & -\frac{T_1}{2} \left(\log(2\pi e) + \log(\hat{\sigma}_z^2) \right) - \sum_{o_t=1} \log(z_t) \\ & + \sum_{o_t=1} \log(\hat{p}_t) + \sum_{o_t=0} \log(1-\hat{p}_t) \end{aligned}, \qquad (5)$$

θ is the vector of the parameters,

σ_z^2 is the variance of the residuals of demand sizes,

$\mathbf{Y}$ is the vector of actual values,

T_1 number of observations of non-zero demand.

Model selection

The selection can be done using AIC, AICc, BIC etc.

e.g. calculating AIC:

$$AIC = 2k - 2\ell(\theta, \sigma^2 | \mathbf{Y}), \tag{6}$$

where k is the number of parameters in the model,

Model selection

We need to know the number of parameters.

It is easy for the models for z_t:

k = smoothing parameters + initial states + 1 + i.

i is equal to one if $o_t \neq 1$ for any t.

This is because we split the data in two parts:

1. z_t – demand sizes;
2. o_t – demand occurrences.

We estimate $\hat{p}_t$ on a separate time series and use it in likelihood.

Experiments

Data

- WF Wholesale data (Johnston et al., 1999);
- Daily data with working days only;
- One year – 248 observations;
- 120 branches, around 600 SKUs;
- Some series have negative values;
- Excluded series with less than 5 non-zero observations;
- Excluded data with no variability;
- We aggregated SKU for all branches to have non-intermittent data;
- Overall – 10221 time series.

Contestants

- iETS(Z,Z,N);
- ETS(A,N,N);
- Croston;
- TSB;
- Naive;
- Zeroes.

`es()` function from `smooth` package for R (from CRAN) for all.

Error measures

- sMSE - Mean Squared Error;
- MREb - Mean Root Error bias;
- sPIS - Periods-in-stock;
- sCE - Cumulative Error;
- PLS - Prediction Likelihood Score;
- Prediction intervals coverage (distance from 95%).

Other settings

- Horizon of 20 days (one month);
- Fixed origin.

Results

Model	MREb	sMSE	sPIS	sCE	PLS	PI
iETS(ZZN)	**-0.640**	0.550	-5.014	-0.535	**-15.621**	0.070
ETS(ANN)	-0.686	**0.547**	**-2.141**	**-0.263**	-114.62	**0.040**
Croston	-0.746	0.556	8.616	0.761	-19.627	0.072
TSB	-0.677	**0.547**	-2.502	-0.298	-18.033	0.120
Naive	0.837	0.761	-2.853	-0.331	-95.841	0.049
Zeroes	0.979	0.578	-21.746	-2.131	-113.343	0.040

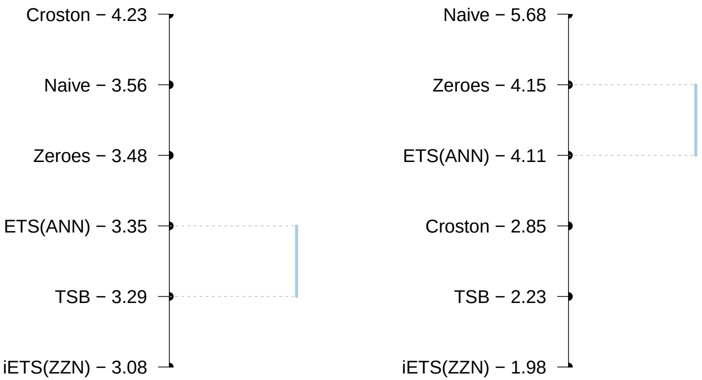
Table: Mean Error measures.

Results

Model	MREb	sMSE	sPIS	sCE	PLS	PI
iETS(ZZN)	**-0.753**	0.018	3.343	**0.241**	**-7.338**	0.050
ETS(ANN)	-0.787	0.020	5.373	0.478	-42.811	0.050
Croston	-0.850	0.031	11.41	1.026	-8.120	0.050
TSB	-0.781	0.019	5.018	0.466	-7.713	0.050
Naive	1.000	0.020	**-2.131**	-0.345	-50.179	0.050
Zeroes	1.000	**0.015**	-4.100	-0.571	-43.038	0.050

Table: Median Error measures.

Nemenyi test (Demšar, 2006) on sMSE and PLS



Nemenyi test on absolute sPIS and Coverage

Conclusions

Conclusions

- Connection between intermittent and conventional models;
- We can use one model for wide variety of series;
- Categorisation based on modelling approach;
- Good results on real data;
- But the experiment needs to be extended.

LCF

Future experiments

- Add Bootstrap to the list of competitors;
- Another dataset (more heterogeneous).

Finale

Thank you for your attention!

Ivan Svetunkov

i.svetunkov@lancaster.ac.uk

Lancaster University Management School | Lancaster Centre for Forecasting

Croston, J. D., sep 1972. Forecasting and Stock Control for Intermittent Demands. Operational Research Quarterly (1970-1977) 23 (3), 289.
URL
http://www.jstor.org/stable/3007885?origin=crossref

Demšar, J., 2006. Statistical Comparisons of Classifiers over Multiple Data Sets. Journal of Machine Learning Research 7, 1–30.

Hyndman, R. J., Koehler, A. B., Ord, J. K., Snyder, R. D., 2008. Forecasting with Exponential Smoothing. Springer Series in Statistics. Springer Berlin Heidelberg, Berlin, Heidelberg.
URL
http://link.springer.com/10.1007/978-3-540-71918-2

Johnston, F. R., Boylan, J. E., jan 1996. Forecasting for Items with Intermittent Demand. Journal of the Operational Research Society 47 (1), 113–121.
URL http://link.springer.com/10.1057/jors.1996.10

Johnston, F. R., Boylan, J. E., Meadows, M., Shale, E., dec 1999. Some Properties of a Simple Moving Average when Applied to Forecasting a Time Series. The Journal of the Operational Research Society 50 (12), 1267–1271.
URL http://wrap.warwick.ac.uk/13847/http://www.jstor.org/stable/3010636?origin=crossref

Petropoulos, F., Kourentzes, N., jun 2015. Forecast combinations for intermittent demand. Journal of the Operational Research Society 66 (6), 914–924.
URL http://dx.doi.org/10.1057/jors.2014.62{%}7B{%}25{%}7D5Cn10.1057/jors.2014.62http://link.springer.com/10.1057/jors.2014.62

Prestwich, S., Tarim, S., Rossi, R., Hnich, B., oct 2014. Forecasting intermittent demand by hyperbolic-exponential smoothing. International Journal of Forecasting 30 (4), 928–933.
URL http://dx.doi.org/10.1016/j.ijforecast.2014.01.

006http://linkinghub.elsevier.com/retrieve/pii/S0169207014000491

Syntetos, A. A., Boylan, J. E., apr 2005. The accuracy of intermittent demand estimates. International Journal of Forecasting 21 (2), 303–314.
URL http://linkinghub.elsevier.com/retrieve/pii/S0925527310002306http://linkinghub.elsevier.com/retrieve/pii/S0169207004000792

Syntetos, A. A., Boylan, J. E., Croston, J. D., 2005. On the categorization of demand patterns. Journal of the Operational Research Society 56 (5), 495–503.

Teunter, R. H., Syntetos, A. A., Babai, M. Z., nov 2011. Intermittent demand: Linking forecasting to inventory obsolescence. European Journal of Operational Research 214 (3), 606–615.
URL http://linkinghub.elsevier.com/retrieve/pii/S0377221711004437