Why do zeroes happen?
A model-based approach for demand classification

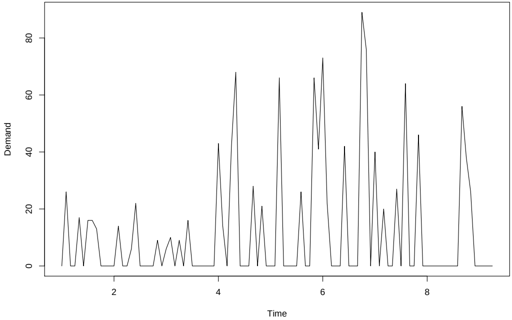
Ivan Svetunkov and Anna Sroginis

Centre for Marketing Analytics and Forecasting, Lancaster University, UK

13th March 2025

Marketing Analytics and Forecasting | Lancaster University Management School

What is intermittent demand?



Intermittent demand is the demand that happens at **irregular frequency** (Svetunkov and Boylan, 2023).

How do you know that the demand is intermittent?

Just set a threshold like 30% and if the number of "zeroes" exceeds this threshold then declare it to be an intermittent demand series. For guidelines to deal with "unusual demands" rather than believing them and Level Shifts (n.b. A level Shift is not a time trend) . Also since intermittent demand can yield rates that are auto-regressive (i.e autocorrelated) models like the Poisson Model or the Croston approach are of limited value. Please see the discussion Please see my comments in How to forecast based on aggregated data over irregular intervals? regarding this.

Share Cite Improve this answer Follow

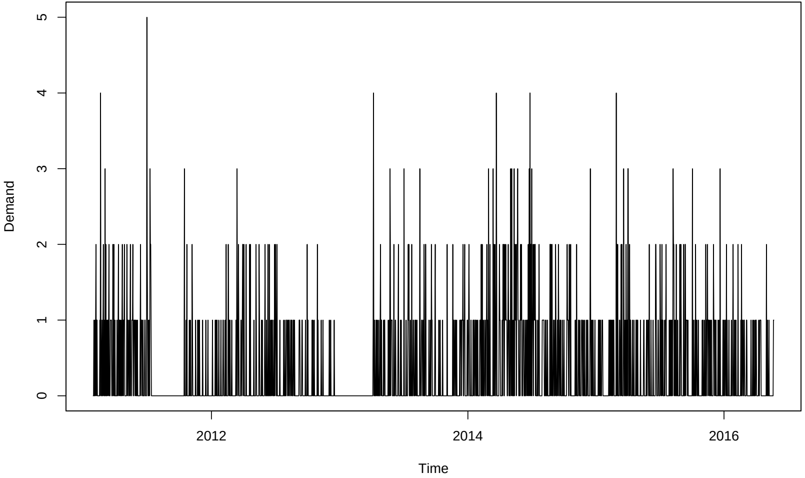
edited Apr 13, 2017 at 12:44
Community Bot
1

answered Mar 8, 2012 at 21:26
IrishStat
30k 5 36 60

30% could be just stockouts.

No good universal rule.

Why do zeroes happen?



Why do zeroes happen?

Several possible reasons:

- Recording started later;
- We stopped selling the product;
- We do not sell the product this time of year;
- Stockouts (product not available);
- Technical issues;
- Other reasons?

In some of these cases, the demand is not intermittent

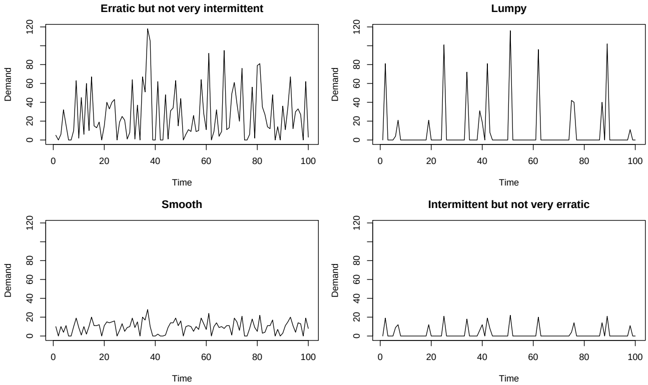
Why should we care?

- We need to forecast demand, not sales (where possible).
- Intermittent demand should be treated differently than the regular one
- If you do it correctly, you might gain in accuracy.
    - e.g. Svetunkov and Boylan (2023) the intermittent demand ETS performed better than the conventional ETS, Poisson and Negative Binomial models.

How can we identify intermittent demand?

Intermittent demand types

Syntetos et al. (2005) developed intermittent demand classification based on coefficient of variation and average demand intervals:



Intermittent demand types

The classification was originally developed to select between Croston (Croston, 1972) and SBA (Syntetos and Boylan, 2005).

It assumes that the demand is intermittent.

Recommendation was to use Croston for the *smooth intermittent demand* and use SBA for all the others.

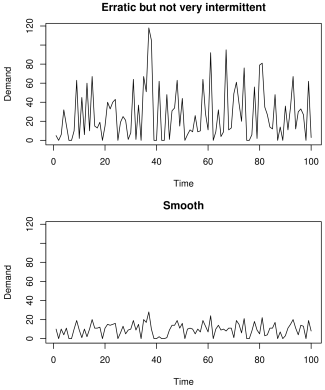
Kostenko and Hyndman (2006) refined the classification, showing that the change from Croston to SBA is not linear.

Petropoulos and Kourentzes (2015) extended it by adding SES for the regular demand.

Why do zeroes happen?

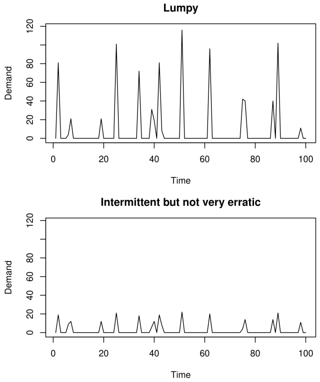
The context is important.

- “Erratic but not very intermittent”:
  - Are zeroes really random?
  - Stockouts maybe?
- “Smooth”:
  - Zeroes happen because not many people buy the product;



Why do zeroes happen?

- "Lumpy":
  - A wholesaler?
  - Sales could be due to external events.
  - What about seasonality?
- "Intermittent but not very erratic":
  - people buy less frequently than in case of the "smooth";
  - spare parts?



Intermittent demand types

Syntetos and Boylan (2006) showed that SES/SMA perform well on intermittent demand.

We also now have lots of advanced approaches for forecasting.

Selecting between SBA/Croston might not be relevant anymore.

e.g. Rožanec et al. (2022) developed ML approaches separately for regular and intermittent demand.

The paper has fundamental issues, but shows that the SBC does not help.

Motivation

We need to have a new look at what intermittent demand is.

The classification needs to be updated.

Ideally, it should help increasing accuracy irrespective of the used approach (stats/ML).

Demand Types

Why do zeroes happen?

Zeroes happen either:

1. Naturally,
2. Artificially.

Former means, nobody wanted to buy the product.

Latter means, sales didn't happen or were not recorded.

(1) is an important part of demand.

Forecasting (2) is pointless.

Demand types

Artificially occurring zeroes need to be treated.

After removing or interpolating them, we can have:

I. Regular demand (no zeroes);
II. Intermittent demand (some naturally occurring zeroes).

Intermittent demand in turn can be:

1. Smooth (e.g. retailer demand),
2. Lumpy (e.g. wholesaler demand).

Demand sizes can be one of two types:

1. Count, e.g. number of engines;
2. Fractional, e.g. litres of milk.

Demand types

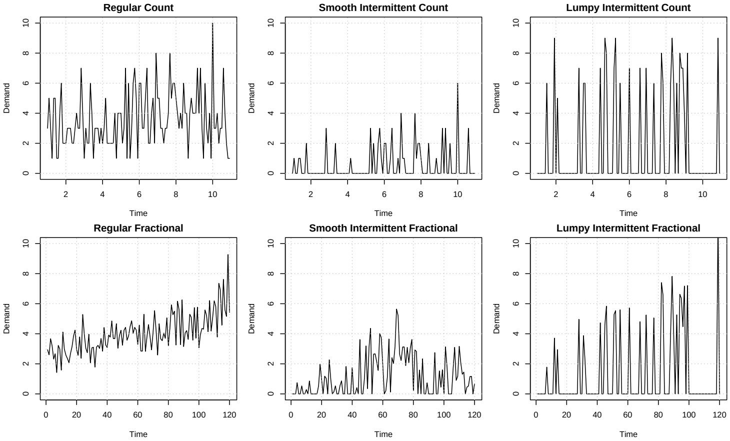


Figure: Examples of demand in different categories

AID

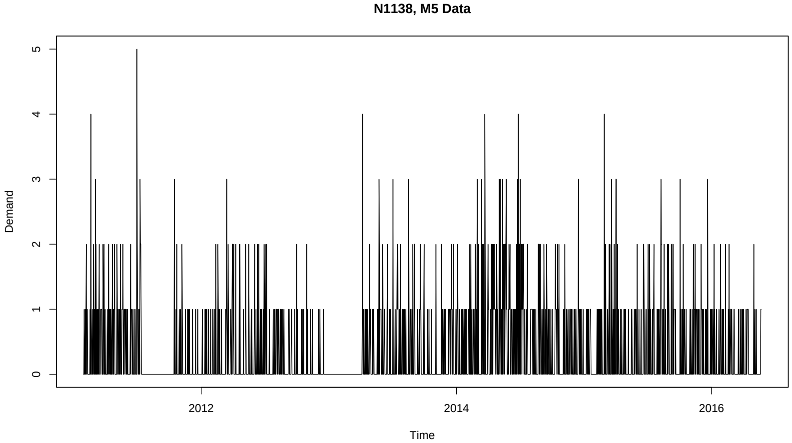
Automated Identification of Demand (AID)

How do we identify demand type?

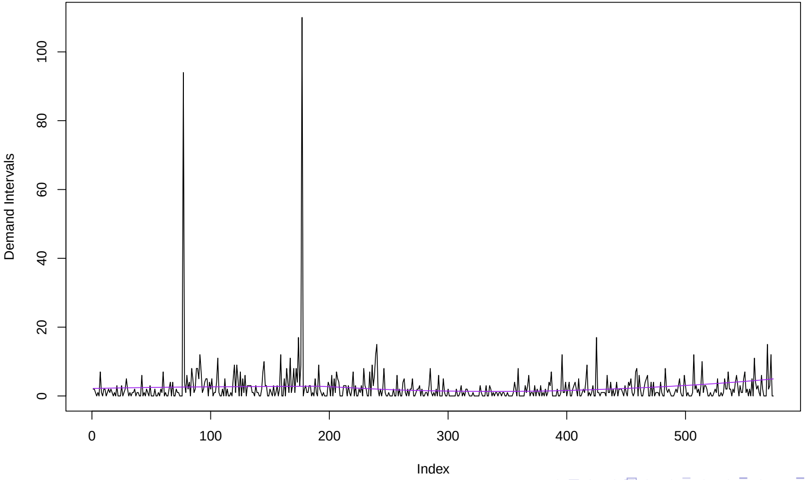
First step – identify stockouts.

Here, we focus on demand intervals...

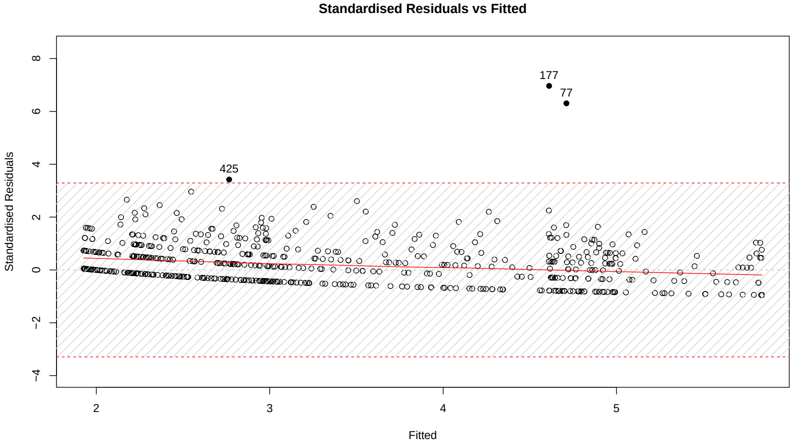
Stockouts identification: example



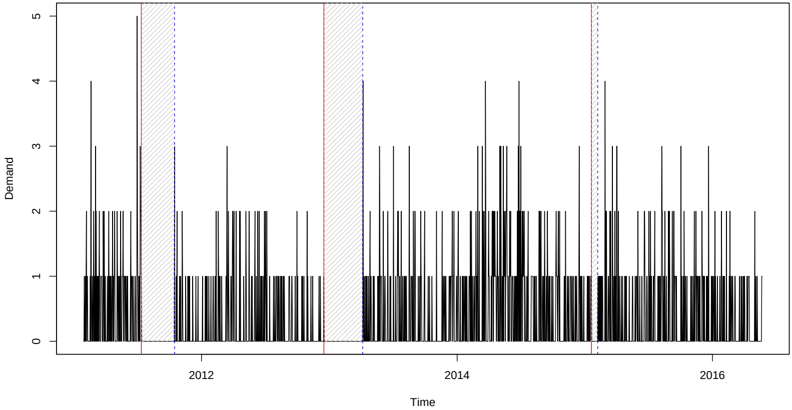
Stockouts identification: demand intervals



Stockouts identification: Outliers



Stockouts identification: Result



Automated Identification of Demand (AID)

After removing stockouts, we check whether there are any zeroes left.

If no, this must be regular demand.

If yes, it is probably intermittent (only naturally occurring zeroes left).

To continue identification, we smooth (Friedman, 1984):

- original series $\hat{y}_t$,
- demand sizes $\hat{z}_t$,
- probability of occurrence $\hat{p}_t$.

Automated Identification of Demand (AID)

We then fit several models with those smoothed series and select the one with the lowest AICc:

I. **Regular Fractional** – the model applied to the data itself, $y_t \sim \mathcal{N}(\beta_0 + \beta_1 \hat{y}_t, \sigma_y^2)$;

II. **Smooth Intermittent Fractional** – $y_t \sim \mathrm{rect}\mathcal{N}(\beta_0 + \beta_1 \hat{y}_t, \sigma_y^2)$ – this model uses the Rectified Normal distribution;

III. **Lumpy Intermittent Fractional** – the mixture distribution model: $y_t = o_t z_t$, where for demand sizes, $z_t \sim \mathcal{N}(\log \beta_0 + \beta_1 \hat{z}_t, \sigma_z^2)$ and for the probability of occurrence, $o_t \sim \mathrm{Bernoulli}(\beta_0 + \beta_1 \hat{p}_t)$;

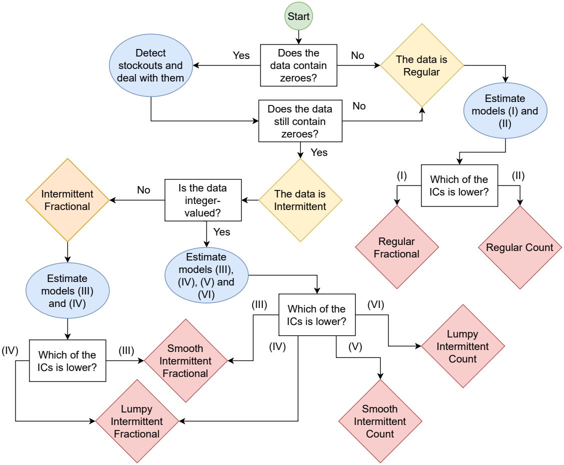
Automated Identification of Demand (AID)

IV. **Regular Count** – Negative Binomial distribution, $y_t \sim \mathcal{NB}(\beta_0 + \beta_1 \hat{y}_t, s_y)$;

V. **Smooth Intermittent Count** – same as (IV), but with zeroes;

VI. **Lumpy Intermittent Count** – the mixture distribution $y_t = o_t z_t$, where demand sizes are $z_t \sim \mathcal{NB}(\beta_0 + \beta_1 \hat{z}_t, s_z)$ and occurrence is $o_t \sim \mathrm{Bernoulli}(\beta_0 + \beta_1 \hat{p}_t)$;

Automated Identification of Demand (AID)



Simulations: Stockouts identification

Stockout identification

Does the stockout detection mechanism work?

Several things that should impact its accuracy:

1. Length of stockouts;
2. Number of stockouts;
3. Number of naturally occurring zeroes in the data;
4. Sample size.

Stockout identification

Our expectation are:

I. The method should be able to detect longer stockouts easier than the shorter ones;
II. The method should find it hard to detect the stockouts when there are more of them in the data;
III. Its power should be inverse to the overall number of zeroes in the data;
IV. Its power should be proportional to the sample size.

Stockout identification

Parameters	Scenario I	Scenario II	Scenario III	Scenario IV
Sample size	100	100	100	30 – 1000
Probability of occurrence	0.8	0.8	0.1 – 0.9	0.8
Number of stockouts	1	1 – 10	5	5
Length of stockouts	3 – 10	5	5	5

Table: The settings for the four scenarios in the first simulation experiment to track stockouts.

Stockout identification

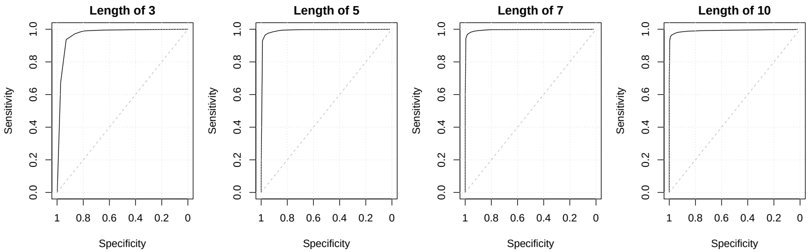


Figure: Scenario 1: changing the length of stockouts.

AUC values: 0.966, 0.973, 0.976 and 0.972

Stockout identification

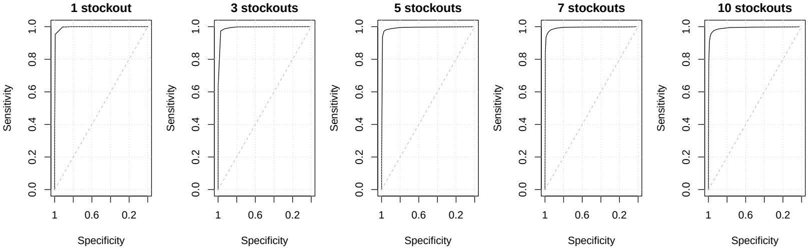


Figure: Scenario 2: changing the number of stockouts.

AUC values: 0.996, 0.977, 0.973, 0.975 and 0.969

Stockout identification

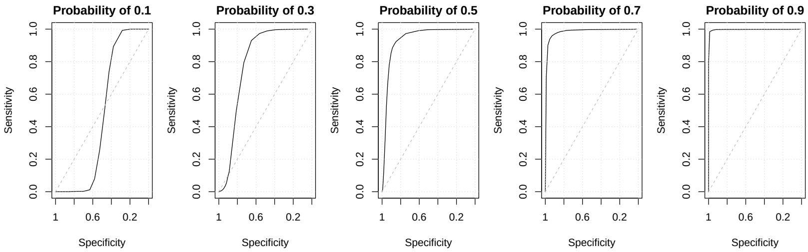


Figure: Scenario 3: changing the probability of occurrence.

AUC values: 0.473, 0.748, 0.909, 0.964 and 0.981

Stockout identification

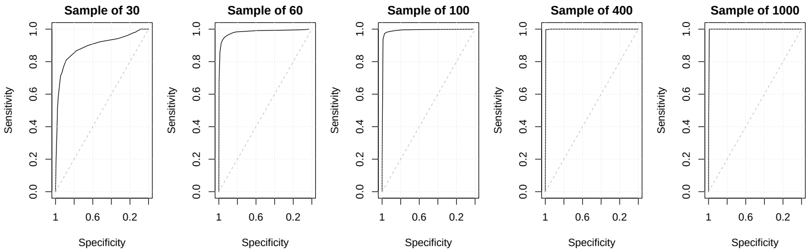


Figure: Scenario 4: varying sample size.

AUC values: 0.894, 0.949, 0.973, 0.994 and 0.996

Summary

The power of the stockouts detection approach is positively related to:

- the sample size,
- the probability of occurrence,
- the length of stockouts,

and negatively related to:

- the amount of stockouts.

Simulations: Demand identification

Demand identification

Create synthetic data using several Data Generating Processes (DGP):

1. **Regular Fractional**: ETS(M,N,N) with the Log-Normal distribution of the residuals;
2. **Smooth Intermittent Fractional**: ETS(A,N,N) with the normal distribution. After generating the data, all negative values were substituted by zeroes. This aligns with the rectified Normal distribution;
3. **Lumpy Intermittent Fractional**: ETS(M,N,N), similar to (1), but after generating the data, random zeroes were introduce (so that 30% of observations are zero);

Demand identification

Create synthetic data using several Data Generating Processes (DGP):

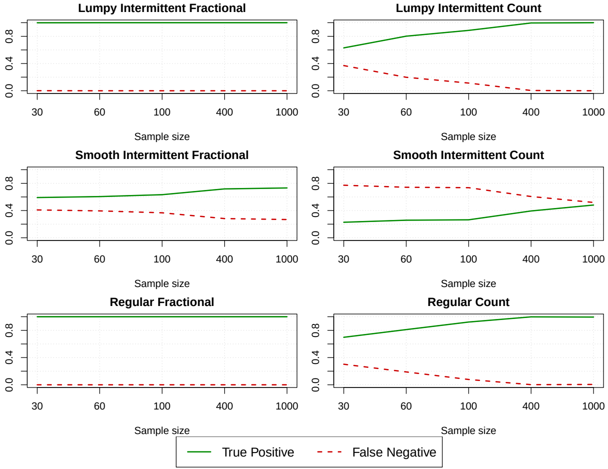
4. **Regular Count**: First, the data was generated using ETS(M,N,N), then it was used in the data generation from the Negative Binomial distribution with size 20 and the mean equal to the ETS(M,N,N) data.;
5. **Smooth Intermittent Count**: Similar to (4), but with lower initial level (5 instead of 10) and lower size (2 instead of 20);
6. **Lumpy Intermittent Count**: Similar to (4), but introducing random zeroes (30% of them).

Demand identification

We generated samples of 30, 60, 100, 400 and 1000 observations.

We applied AID via the `aid()` function from the `greybox` package (Svetunkov, 2024) in R (R Core Team, 2020) and recorded how demand was classified.

Demand identification



Summary

Main issue is the stockout identification.

Regulating the confidence level, changes sensitivity of the approach.

It is sometimes hard to tell the difference between the fractional/count.

Case Study

Case Study: The setting

Demand data of a retailer

342 weekly observations,

Starting from 1st April 2018 and finishing on 4th November 2024

Some products having shorter histories than the others.

The dataset contained 3 shops with overall 31018 products.

We used LightGBM (Ke et al., 2017) using the `lightgbm` package in R (Shi et al., 2024).

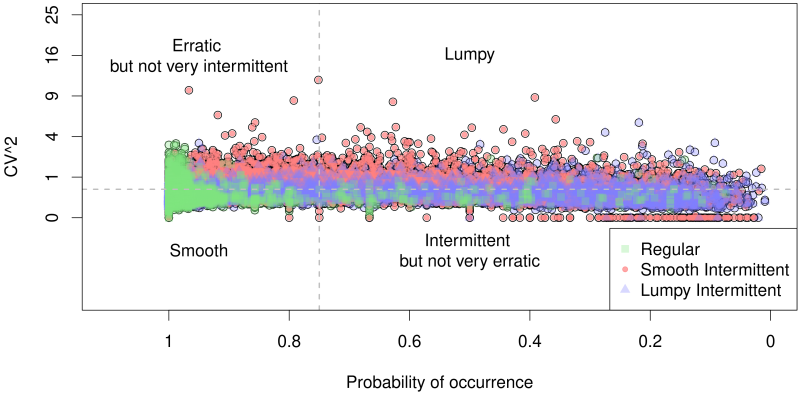
Case Study: Results of the categorisation

Demand sizes type	Regular	Smooth Intermittent	Lumpy Intermittent	
Count	Regular Count (5652)	Smooth Intermittent Count (19128)	Lumpy Intermittent Count (2753)	Count (27533)
Fractional	Regular Fractional (1115)	Smooth Intermittent Fractional (1342)	Lumpy Intermittent Fractional (1016)	Fractional (3473)
	Regular (6767)	Smooth Intermittent (20470)	Lumpy Intermittent (3769)	Total (31006)

Demand intervals

Figure: Demand classification for the retail company data.

Case Study: AID vs SBC



Case Study: Features

1. Promotions – dummy variables, indicating when an item was on promotion;
2. Holidays – categorical variable, denoting holidays;
3. Events – categorical variable, denoting special events for specific dates;
4. Covid – binary variable capturing the effect of covid on sales;
5. Stockouts – dummy variable, showing when stockouts happened according to our approach with the confidence level of 0.999;

Case Study: Features

6. SmoothBasic – smoothed original series. The smoothing was done using Friedman's Super Smoother (Friedman, 1984) via the `supsmu()` function from the `stats` package in R (R Core Team, 2020);
7. Smooth – another version of the smoothed series was done by excluding the observations that were detected as stockouts using our approach;
8. SmoothSizes – smoothed demand sizes only, which was an important feature for the mixture model;
9. Probability – smoothed binary demand occurrence variable (th estimate of the probability of occurrence).

Case Study: LightGBMs

We fit several LightGBMs to see whether features/categories help:

A. Conventional – one LightGBM approach applied directly to the full dataset ignoring the stockouts feature and using promotions, holidays, events, covid and the SmoothBasic;

B. Full – similar to (A), but with the stockouts dummy variable and feature Smooth instead of SmoothBasic;

C. Mixture – two approaches for demand occurrence and demand sizes. Their forecasts are multiplied.
   - The LightGBM 1 focused on predicting the probability of occurrence. Features from (B), but with SmoothSizes and Probability.
   - The LightGBM 2 focused on predicting the demand sizes. Features from (B), but with SmoothSizes instead of Smooth.

Case Study: LightGBMs

More:

D. Category – three LightGBMs, one applied to the data which was flagged as "Regular" in the manner similar to (A), and the other two applied to the data flagged as "Intermittent" in the manner similar to (C);

E. Local Level – just a smoothed series with a straight line for the holdout (similar to SMA).

Case Study: Expectations

What we want to see:

i. The effect of stockout detection mechanism on the accuracy (comparing performance of approaches (A) and (B));
ii. The impact of modelling intermittent demand via the demand occurrence and demand sizes parts on the accuracy (comparing approaches (B) and (C));
iii. The usefulness of the proposed demand classification by comparing (D) with (B) and (C).

We measure performance of approaches in terms of Root Mean Squared Scaled Error (RMSSE) from Makridakis et al. (2022), originally motivated by Athanasopoulos and Kourentzes (2023).

Case Study: Results

	min	Q1	median	mean	Q3	max
Conventional	0.0008	0.4248	0.6950	0.9768	1.1031	125.4219
Full	0.0065	0.4111	0.6737	0.9260	1.0763	**60.0927**
Mixture	**0.0000**	**0.2222**	**0.4629**	0.6638	0.8305	125.6691
Category	**0.0000**	0.2267	0.4690	**0.6612**	**0.8213**	64.7943
Local Level	**0.0000**	0.4255	0.6898	0.9625	1.0984	78.3027

Table: Performance of several LightGBM approaches with different features in terms of RMSSE on the retailer company data. Q1 and Q3 are the first and third quartiles respectively.

Case Study: Results

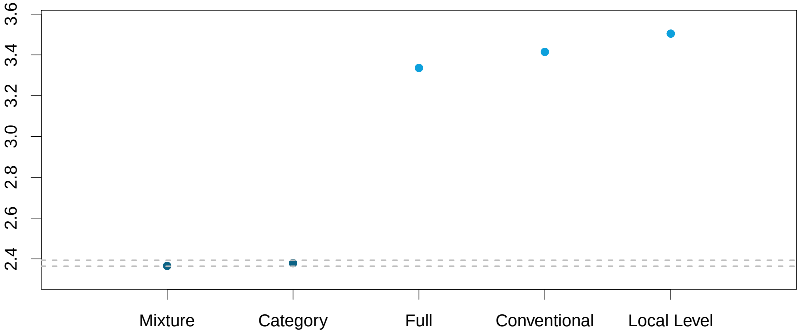


Figure: Nemenyi test for the five LightGBM approaches conducted with the 5% significance level.

Conclusions

Conclusions

- The existing demand classification approaches are outdated;
- Not every demand that has zeroes is intermittent;
- Zeroes can happen for a variety of reasons;
- Artificially occurring ones need to be treated;
- Splitting demand between regular and intermittent becomes straightforward;
- Intermittent can be categorised into smooth and lumpy;
- We also have count/fractional categories;
- But none of that matters, when it comes to forecasting!

Thank you for your attention!

Ivan Svetunkov

i.svetunkov@lancaster.ac.uk

https://www.openforecast.org/

Marketing Analytics and Forecasting | Lancaster University Management School

References I

George Athanasopoulos and Nikolaos Kourentzes. On the evaluation of hierarchical forecasts. *International Journal of Forecasting*, 39(04):1502–1511, 2023. ISSN 01692070. doi: 10.1016/j.ijforecast.2022.08.003.

J D Croston. Forecasting and stock control for intermittent demands. *Operational Research Quarterly (1970-1977)*, 23:289, 9 1972. ISSN 00303623. doi: 10.2307/3007885.

Jerome H Friedman. *A variable span smoother*. Laboratory for Computational Statistics, Department of Statistics, Stanford University, 1984.

References II

Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, volume 30, pages 3146–3154. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf.

A V Kostenko and R J Hyndman. A note on the categorization of demand patterns. *Journal of the Operational Research Society*, 57:1256–1257, 10 2006. ISSN 0160-5682. doi: 10.1057/palgrave.jors.2602211.

References III

Spyros Makridakis, Evangelos Spiliotis, and Vassilios Assimakopoulos. M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*, 38: 1346–1364, 10 2022. ISSN 01692070. doi: 10.1016/j.ijforecast.2021.11.013.

Fotios Petropoulos and Nikolaos Kourentzes. Forecast combinations for intermittent demand. *Journal of the Operational Research Society*, 66:914–924, 6 2015. ISSN 0160-5682. doi: 10.1057/jors.2014.62.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL https://www.R-project.org/.

References IV

Jože Martin Rožanec, Blaž Fortuna, and Dunja Mladenić. Reframing Demand Forecasting: A Two-Fold Approach for Lumpy and Intermittent Demand. *Sustainability*, 14(15):9295, January 2022. ISSN 2071-1050. doi: 10.3390/su14159295.

Yu Shi, Guolin Ke, Damien Soukhavong, James Lamb, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, Tie-Yan Liu, Nikita Titov, and David Cortes. *lightgbm: Light Gradient Boosting Machine*, 2024. URL https://CRAN.R-project.org/package=lightgbm. R package version 4.5.0.

References V

Ivan Svetunkov. *greybox: Toolbox for Model Building and Forecasting*, 2024. URL https://CRAN.R-project.org/package=greybox. R package version 2.0.4.

Ivan Svetunkov and John E. Boylan. iets: State space model for intermittent demand forecasting. *International Journal of Production Economics*, 265:109013, 2023. ISSN 09255273. doi: 10.1016/j.ijpe.2023.109013.

Aris A. Syntetos and John E. Boylan. The accuracy of intermittent demand estimates. *International Journal of Forecasting*, 21: 303–314, 4 2005. ISSN 01692070. doi: 10.1016/j.ijforecast.2004.10.001.

References VI

Aris A. Syntetos and John E. Boylan. On the stock control performance of intermittent demand estimators. *International Journal of Production Economics*, 103:36–47, 2006. ISSN 09255273. doi: 10.1016/j.ijpe.2005.04.004.

Aris A. Syntetos, John E. Boylan, and J D Croston. On the categorization of demand patterns. *Journal of the Operational Research Society*, 56:495–503, 2005. ISSN 0160-5682. doi: 10.1057/palgrave.jors.2601841.