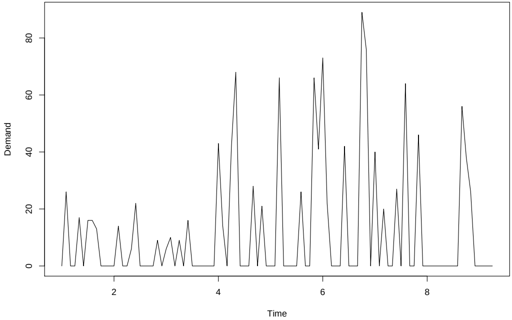
Why do zeroes happen? A model-based view on demand classification

Ivan Svetunkov and Anna Sroginis

Centre for Marketing Analytics and Forecasting, Lancaster University, UK

1st May 2025

What is intermittent demand?



Intermittent demand is the demand that occurs 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?

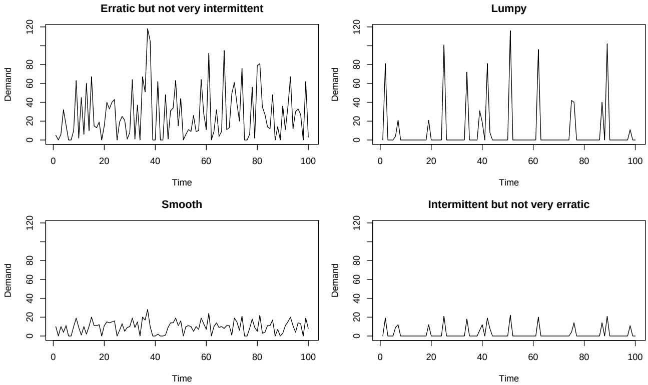
Several possible reasons:

- 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 cases, the demand is not intermittent

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.

Motivation

SBC is often misused.

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

We also now have lots of advanced approaches for forecasting.

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

Two types of zeroes

Zeroes happen either:

1. Naturally,
2. Artificially.

Former means, nobody wanted to buy the product.

Latter means, nobody could buy the product.

(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).

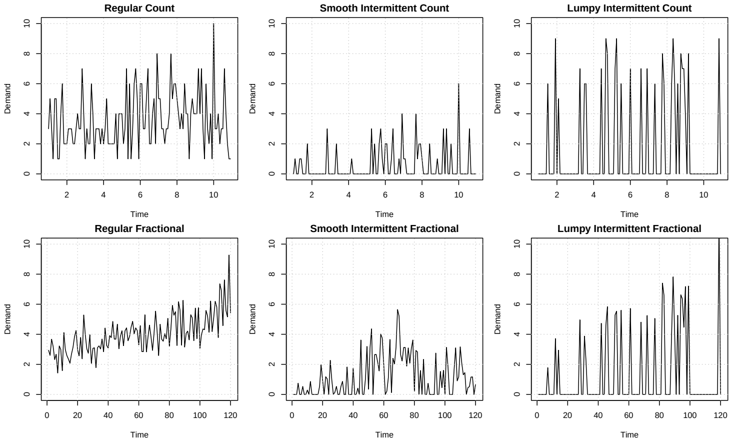
Intermittent demand in turn can be:

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



AID

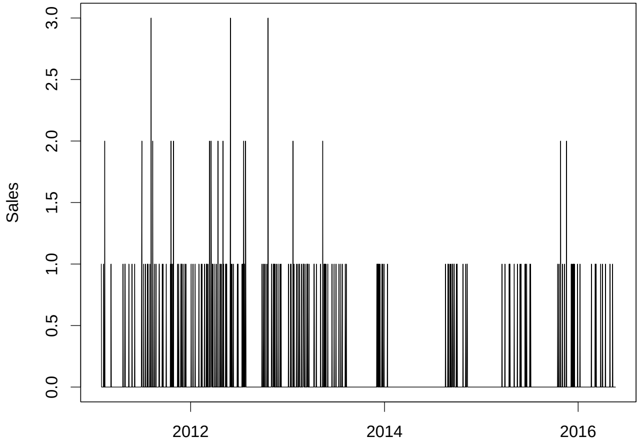
Automated Identification of Demand (AID)

How do we identify demand type?

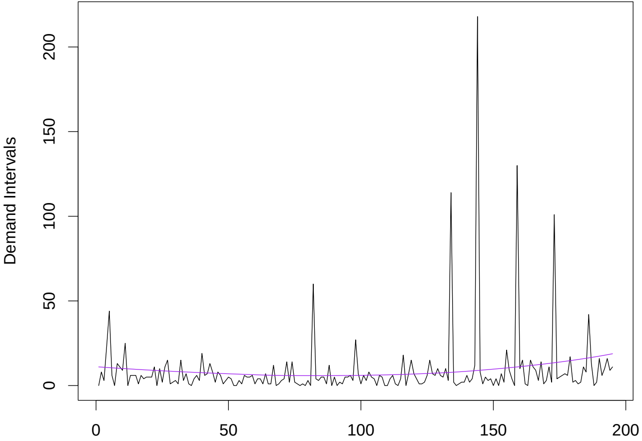
First step – identify stockouts.

Here, we focus on demand intervals...

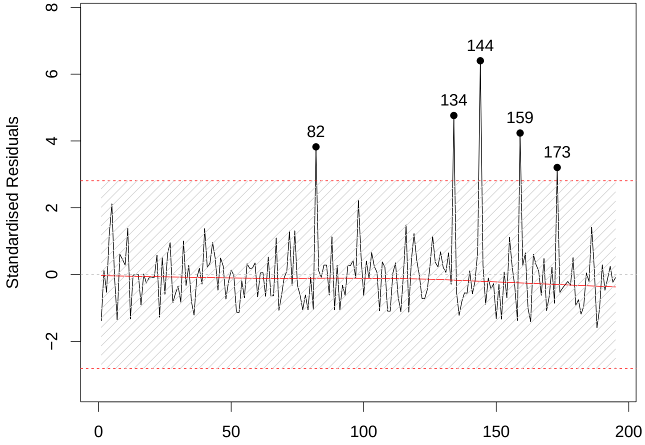
Stockouts identification: example



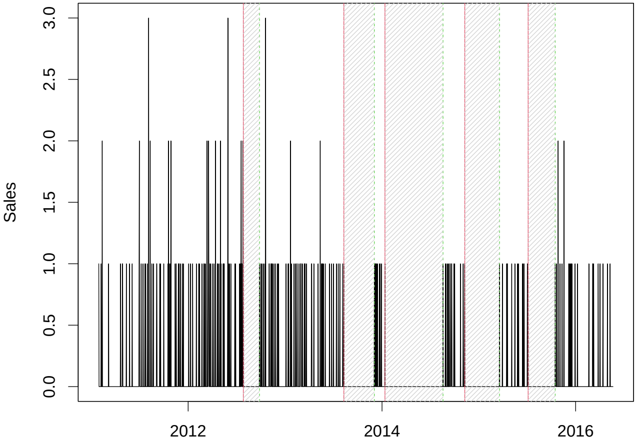
Stockouts identification: demand intervals



Stockouts identification: Outliers



Stockouts identification: Result



Automated Identification of Demand (AID)

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

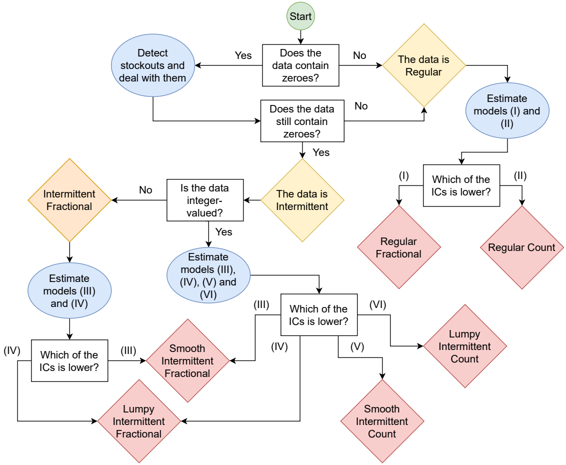
If no, this must be regular demand.

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)



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 removed data that had only zeroes in the holdout)

Case Study: Distribution of stockouts lengths

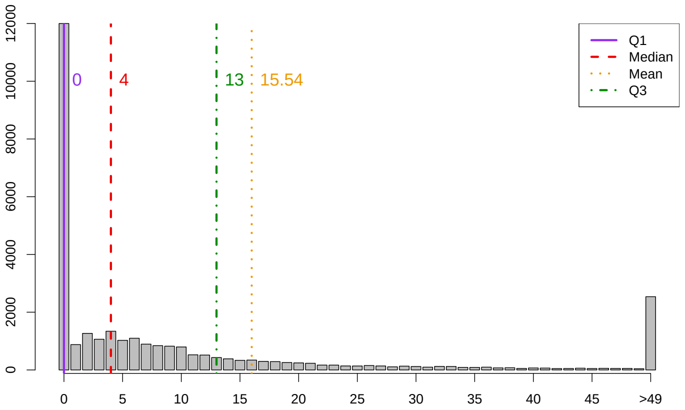


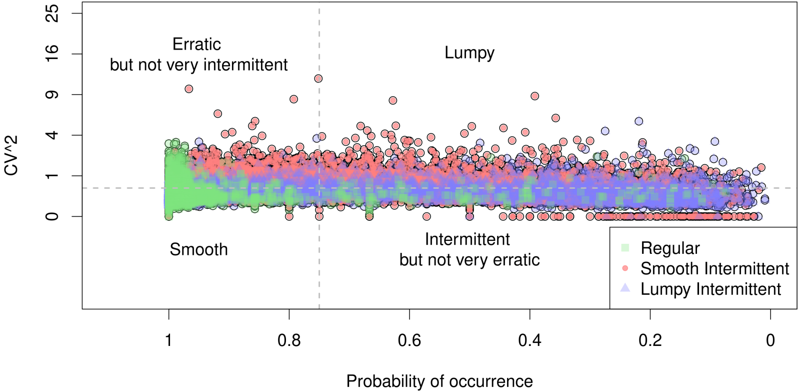
Figure: Distribution of number of stockouts per series for the retail data.

Case Study: Results of the categorisation

	Regular	Smooth Intermittent	Lumpy Intermittent	Overall
Count	5652	19128	2753	27533
Fractional	1115	1342	1016	3473
Overall	6767	20470	3769	31006

Table: 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;

Case Study: Features

5. Stockouts – dummy variable, showing when stockouts happened according to our approach with the confidence level of 0.999;
6. SmoothSales – smoothed original series. The smoothing was done using Friedman's Super Smoother (Friedman, 1984);
7. SmoothDemand – smoothed "Demand" after removing stockouts;
8. SmoothSizes – smoothed demand sizes only;
9. Probability – smoothed binary demand occurrence variable (the estimate of the probability of occurrence).

Case Study: LightGBMs

We fit several LightGBMs (Ke et al., 2017; Shi et al., 2024):

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

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

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

Case Study: LightGBMs

More:

D. CategoryPartial – 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. CategoryFull – Same as (D), but applied to Regular/Intermittent Smooth/Intermittent Lumpy.

And then Pooled Regression and just smoothed series (local level).

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: LightGBM Results

	min	Q1	median	mean	Q3	max
Conventional	0.0035	0.4154	0.6814	0.9535	1.0975	65.2722
Full	0.0030	0.4130	0.6735	0.9247	1.0738	**60.0936**
Mixture	0.0000	**0.2204**	**0.4642**	0.6646	0.8289	125.4443
Category Partial	0.0000	0.2235	0.4677	**0.6601**	**0.8239**	68.3509
Category Full	0.0000	0.2221	0.4668	0.6637	0.8251	124.0355

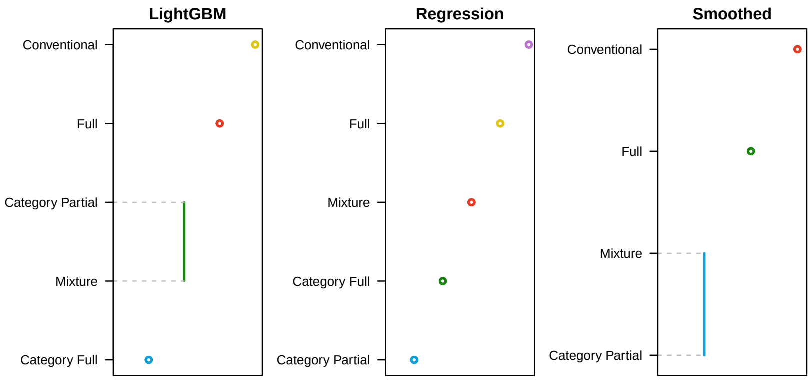
Case Study: Pooled Regression Results

	min	Q1	median	mean	Q3	max
Conventional	0.0002	0.4295	0.6950	0.9620	1.0980	**73.5414**
Full	0.0004	0.4407	0.7090	0.9596	1.0942	73.6920
Mixture	0.0000	**0.3031**	**0.5770**	**0.8422**	**0.9984**	79.6121
Category Partial	0.0000	0.3091	0.5826	0.8526	1.0027	76.5965
Category Full	0.0000	0.3090	0.5829	0.8517	1.0019	78.1593

Case Study: Local Level Results

	min	Q1	median	mean	Q3	max
Conventional	0.0000	0.4255	0.6898	0.9625	1.0985	78.3027
Full	0.0000	0.4277	0.6906	0.9552	1.0954	78.3027
Mixture	0.0000	**0.3043**	**0.5762**	**0.8380**	**0.9973**	**75.5905**
Category Partial	0.0000	0.3107	0.5867	0.8456	1.0051	**75.5905**

Case Study: Results



Conclusions

Conclusions

- Not every demand that has zeroes is intermittent;
- Artificially occurring zeroes need to be treated;
- Splitting demand between regular and intermittent becomes straightforward;
- Intermittent can be categorised into smooth and lumpy;
- Mixture approaches bring the largest value;
- Category-wise split brings further (but small) improvements.

Thank you for your attention!

Ivan Svetunkov

i.svetunkov@lancaster.ac.uk

https://www.openforecast.org/

Marketing Analytics and Forecasting

Lancaster University Management School

Appendix A. 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, $y_t \sim \mathcal{N}(\beta_0 + \beta_1 \hat{y}_t, \sigma^2_y)$;

II. **Smooth Intermittent Fractional** – $y_t \sim \mathrm{rect}\mathcal{N}(\beta_0 + \beta_1 \hat{y}_t, \sigma^2_y)$ – 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, $z_t \sim \mathcal{N}(\log \beta_0 + \beta_1 \hat{z}_t, \sigma^2_z)$ and for the probability of occurrence, $o_t \sim \mathrm{Bernoulli}(\beta_0 + \beta_1 \hat{p}_t)$;

Appendix A. 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 $o_t \sim \mathrm{Bernoulli}(\beta_0 + \beta_1 \hat{p}_t)$;

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.

References IV

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.

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.

References V

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.

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.