Contents lists available at ScienceDirect

Int. J. Production Economics

Another look at estimators for intermittent demand

Fotios Petropoulos [a,*], Nikolaos Kourentzes [b], Konstantinos Nikolopoulos [c]

[a] *Logistics and Operations Management Section, Cardiff Business School, Cardiff University, UK*
[b] *Department of Management Science, Lancaster University Management School, Lancaster University, UK*
[c] *Bangor Business School, Bangor University, UK*

ABSTRACT

In this paper we focus on forecasting for intermittent demand data. We propose a new aggregation framework for intermittent demand forecasting that performs aggregation over the demand volumes, in contrast to the standard framework that employs temporal (over time) aggregation. To achieve this we construct a transformed time series, the inverse intermittent demand series. The new algorithm is expected to work best on erratic and lumpy demand, as a result of the variance reduction of the non-zero demands. The improvement in forecasting performance is empirically demonstrated through an extensive evaluation in more than 8000 time series of two well-researched spare parts data sets from the automotive and defence sectors. Furthermore, a simulation is performed so as to provide a stock-control evaluation. The proposed framework could find popularity among practitioners given its suitability when dealing with clump sizes. As such it could be used in conjunction with existing popular forecasting methods for intermittent demand as an exception handling mechanism when certain types of demand are observed.

1. Introduction

Forecasting for lumpy and erratic demand is challenging. Intermittent demand is characterised by variable demand sizes coupled with irregular demand arrivals, with many observations having zero demand. Such demand patterns are very common in many industrial settings. Given that inventory management and stock control builds on demand forecasting, it is obvious that accurately forecasting irregular demands is strongly linked with optimal inventory levels.

A first systematic approach to deal with intermittent demand data was introduced by Croston (1972). Croston proposed the decomposition of such data into two separate series, corresponding to the non-zero demand sizes and the inter-demand intervals. Each series is extrapolated separately and the ratio of the two forecasts corresponds to Croston's method forecasts. Croston's method deals independently with each type of variance observed in the data: the variance of the non-zero demands and that of the inter-demand intervals. Various alternatives to Croston's method have been proposed in the literature, most notably the Syntetos–Boylan approximation (Syntetos and Boylan, 2005) and the Teunter–Syntetos–Babai method (Teunter et al., 2011). However, all

of them build on the same concept, separating the non-zero demands from the intervals or the probability to have a non-zero demand. Croston's method and its variants attempt to provide an estimate of the demand rate, thus answering the question "what will the mean demand be for every future period?".

An alternative approach to deal with the intermittence of demand comes from Nikolopoulos et al. (2011). They proposed the Aggregate–Disaggregate Intermittent Demand Approach (ADIDA), which uses equally sized time buckets to perform non-overlapping temporal aggregation. ADIDA reduces the variance observed in the intervals. Thus, given an appropriate level of aggregation that removes the intermittence of the data, ADIDA focuses only on forecasting the (aggregated) non-zero demands. One could argue that the improvements in performance offered by the ADIDA framework originate from the reduction (or elimination) of the variance observed in the intervals. If a disaggregation mechanism is not considered at the very last stage of this framework, then ADIDA is addressing the question "how many SKUs will be sold over a pre-specified lead time?".

However, instead of focusing on minimising the variance of the inter-demand intervals and modelling the variance of the demand, one could fix the "demand buckets", thus minimising the variance of the demand, and forecast the respective time-varying number of periods that this fixed demand will occur. We define such an alternative view of the data as Inverse ADIDA. This approach performs non-overlapping temporal aggregation creating equally sized demand buckets, which makes sense from a managerial

---

* Corresponding author.
E-mail addresses: petropoulosf@cardiff.ac.uk (F. Petropoulos),
n.kourentzes@lancaster.ac.uk (N. Kourentzes),
k.nikolopoulos@bangor.ac.uk (K. Nikolopoulos).

viewpoint, as many SKUs are distributed in pallets with pre-fixed quantities. In other words, Inverse ADIDA focuses on answering the question: "when should I re-order a new pallet?".

We empirically examine the new method's performance against standard intermittent demand estimators and the original ADIDA framework. Evaluation is performed both in terms of error metrics and inventory performance, using real and simulated data.

The remaining of this paper is organised as follows. Section 2 provides a quick overview of the main methods proposed for forecasting intermittent demand data. Section 3 gives an alternative view to intermittent demand data and introduces the concept of inverse series and inverse ADIDA. Section 4 outlines the experimental set-up and Section 5 discusses the evaluation of the results. Lastly, Section 6 provides concluding remarks and paths for future research.

2. Background research

2.1. Parametric methods

Simple Exponential Smoothing (SES) is frequently being used as the tool to forecast intermittent demand data. The method is applied directly on the original data. The forecast for a point in the future is calculated as:

$$\hat{y}_t = \hat{y}_{t-1} + \alpha (y_{t-1} - \hat{y}_{t-1}). \tag{1}$$

where α is the smoothing parameter for the level.

However, Croston (1972) pointed out that this method suffers from a decision-point bias, producing very large forecasting errors. Thus, he proposed a decomposition approach so that the updates of the estimate are performed only after a non-zero demand is observed. At the same time, the first differences of the time periods that a non-zero demand has occurred, which measures the inter-demand intervals, are estimated separately. The two estimates are divided to produce the final forecast. Croston's method is the standard method to be used in the industry nowadays, being implemented in many ERP systems and dedicated forecasting software.

Another way to look at Croston's decomposition is that of a non-overlapping temporal aggregation, using time buckets the length of which varies over time, so that: (i) exactly one non-zero demand is included in each time bucket; (ii) the non-zero demand occurs at the end of each bucket; and (iii) intermittence is removed. However, as the length of the time buckets is time-varying, one has to forecast both non-zero demands and inter-demand intervals.

Fig. 1 presents an illustrative example of Croston's decomposition. The x-axis of the main graph represents the time period, which is equivalent to the cumulative value of the inter-demand intervals. The y-axis provides the size of the demand at each period. The different gray shades on the original series represent the time-varying buckets of the non-overlapping temporal aggregation. Note that sometimes the first value of the inter-demand intervals is not taken into account when forecasting, as there is uncertainty with regard to the timing of the previous non-zero demand.

Using SES, each of the two series, the non-zero demand and the inter-demand intervals, is estimated individually:

$$\hat{z}_t = \hat{z}_{t-1} + \alpha_z (z_{t-1} - \hat{z}_{t-1}), \tag{2}$$

$$\hat{p}_t = \hat{p}_{t-1} + \alpha_p (p_{t-1} - \hat{p}_{t-1}). \tag{3}$$

where α_z and α_p are the smoothing parameter for the non-zero demands and the intervals respectively. The final output of Croston's method is simply the division of these estimates:
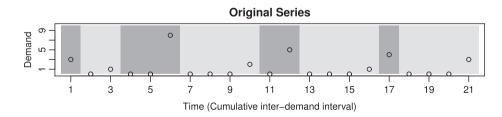
$$\hat{y}_t = \frac{\hat{z}_t}{\hat{p}_t}. \tag{4}$$

The combined output is not an estimate of the actual demand, but a demand rate (Kourentzes, 2013; Petropoulos and Kourentzes, 2015), which if accumulated should be the estimate of the non-zero demand, whenever this occurs.

Syntetos and Boylan (2001) showed that Croston's method is biased. The bias is linked to the value of the smoothing parameter for the inter-demand intervals. Thus, Syntetos and Boylan (2005) suggested an unbiased estimator (Syntetos–Boylan Approximation or SBA) for intermittent demand, by introducing a debiasing factor that is directly multiplied to the output of Croston's method:

$$\hat{y}_t = (1 - \frac{\alpha_p}{2}) \frac{\hat{z}_t}{\hat{p}_t}. \tag{5}$$

Another interesting method for intermittent demand was introduced by Teunter et al. (2011) that has its base back to the original Croston's paper. They suggested that, instead of dividing with the estimate of the inter-demand interval, the forecast of the non-zero demand should be multiplied by the probability to have a non-zero demand. They proposed that, in contrast to the estimates of the demands, the update of the probability estimate to have a non-zero demand should be performed at every period. As
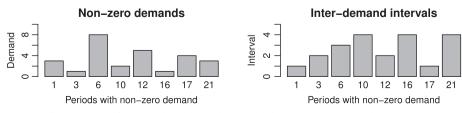

Fig. 1. Example of an intermittent demand series and the respective Croston's decomposition.

a result, the TSB method is useful in the cases where inventory obsolescence should be linked to forecasting.

With regard to the values for the methods' smoothing parameters, the majority of the literature suggest and make use of fixed parameters in the range of 0.05–0.3 (for example see Croston, 1972; Syntetos and Boylan, 2005; Teunter and Duncan, 2009). However, Petropoulos et al. (2013) provided some evidence that optimised smoothing parameters may result to improvement in terms of bias. Kourentzes (2014) revisited the issue of optimal smoothing parameters and initial values proposing two new cost functions, namely the Mean Absolute Rate and the Mean Squared Rate, which demonstrated substantial gains over pre-selected fixed parameters. When using optimal smoothing parameters the performance of the different forecasting methods (Croston's, SBA, TSB) was found to be very similar.

With regard to switching from one method to another, Syntetos et al. (2005), later refined by Kostenko and Hyndman (2006), suggested that the mean inter-demand interval (p) and the squared coefficient of variation of the non-zero demands (v) should be used in order to select when to choose between Croston's method and SBA. Petropoulos and Kourentzes (2015) revised this classification scheme to include the cases where intermittence is not present, where they suggest the use of SES. This new classification scheme will be hereafter referred to as PK.

2.2. Temporal aggregation

A promising approach to deal with the intermittence in demand of slow moving items is the non-overlapping temporal aggregation. The frequency transformation of the original series to lower ones will reduce or even remove the relative number of zero demands per period.

Willemain et al. (1994) were the first to empirically examine the impact of temporal aggregation for intermittent demand. Limited to only a few time series, they showed that improvements in forecast accuracy can be reached by converting daily data to weekly ones. Nikolopoulos et al. (2011) confirmed this result by examining a much larger data set and different disaggregation strategies. More specifically, they proposed the Aggregate–Disaggregate Intermittent Demand Approach (ADIDA), which focuses on transforming the data in a single lower (compared to the original data) frequency and disaggregating the respective forecasts to produce forecasts at the original frequency.

A potential problem with regard to temporal aggregation is the selection of an 'optimal' aggregation level. This problem has been previously examined both empirically, using various in-sample criteria (Spithourakis et al., 2011), and analytically (Rostami-Tabar et al., 2013, 2014). More recently, Petropoulos and Kourentzes (2015) showed that the combined use of multiple aggregation levels results in superior forecasting performance, thus tackling the problem of selecting a single aggregation level. The same insight holds for the case of non-stationary fast moving series (Kourentzes et al., 2014; Kourentzes and Petropoulos, 2016), where temporal aggregation is acting as a filter for high frequency components (seasonality) while enhancing low frequency ones (such as long-term trend). Lastly, Nikolopoulos et al. (2011) suggest that a managerially driven aggregation level that makes sense in a practical inventory settings would be directly linked with the lead time (plus the review period).

As mentioned, temporal aggregation reduces or even removes intermittence. Considering Croston's decomposition at the aggregated series, the variance of the inter-demand intervals is lower compared to the original series, becoming zero for series where intermittence is completely removed. As a result, one can focus solely on estimating the variance of the demands. Despite the added complexity due to demand aggregation (and possibly forecast disaggregation), we argue that non-overlapping temporal aggregation simplifies the problem of forecasting for intermittent demand by tackling the problem of accurately estimating the variance of the inter-demand intervals.

It is worth mentioning that the PK classification becomes particularly relevant to the case of temporally aggregated intermittent demand series, as intermittence may be removed in some aggregation levels.

Assuming an aggregation level equal to 3 periods, Fig. 2 presents the aggregated view of the data presented in Fig. 1, along with the respective sub-series of the non-zero demands and the inter-demand intervals as derived by Croston's decomposition.

3. Another view at intermittent demand

3.1. Inverting the series

Considering the two sub-series that have been derived from Croston's decomposition approach, the non-zero demands and the inter-demand intervals, one may recreate the original series by following the reverse procedure. In other words, the original intermittent demand series and the two sub-series contain the same information. The values of the cumulative inter-demand intervals series are the points in time where a non-zero demand has occurred. The rest of the periods have simply zero demand. This process will lead us back to the original series, which is represented as the volume of demand per period in time, or the volume of demand per the cumulative inter-demand interval.

Instead of considering the cumulative values of the inter-demand intervals, one could calculate the cumulative values of the demand. Then, by linking the values of the inter-demand intervals with the cumulative demand, one new series may be defined: the *inverse intermittent demand series*. This series provides the inter-demand intervals per cumulative demand. Each non-zero value represents the number of periods since the last non-zero demand, while the distance between two non-zero values suggests the volume of demand that occurred for the latter of the two instances.

To illustrate how the inverse of an intermittent demand series can be calculated, consider the data presented in Fig. 1. The decomposition by Croston's method is presented on the bottom panel of that figure. The time periods of non-zero demand occurrences, the values of these non-zero demands and the respective inter-demand intervals are presented for convenience in columns 1, 2 and 3 of Table 1. Column 1 has the cumulative summation of the inter-demand intervals. We also calculate at column 4 the cumulative summation of the non-zero demands.

Columns 1 and 2 constitute the original series, where periods with zero demand have been excluded. In other words, columns 1 and 2 are represented as the values of x-axis and y-axis respectively in Fig. 1. Columns 3 and 4 constitute values of the inverted series, where cumulative demands with zero inter-demand intervals have been excluded. For example, the cumulative demand of nine (9) corresponds to zero (0) inter-demand interval. An interpretation of this is that cumulative demand of nine (9) has never been recorded at the end of a period in the original time frequency. Fig. 3 provides the resulting inverse series. Note that columns 3 and 4 are represented as the values of y-axis and x-axis respectively in Fig. 3.

By forecasting the original series, we end up with an estimate of the demand per each period (or the demand rate). The cumulative demand rate will give an estimate of the demand over a period, which may be the lead-time. On the contrary, by extrapolating the inverse series, we estimate the periods that will pass before a single unit is demanded. The cumulative forecast of the
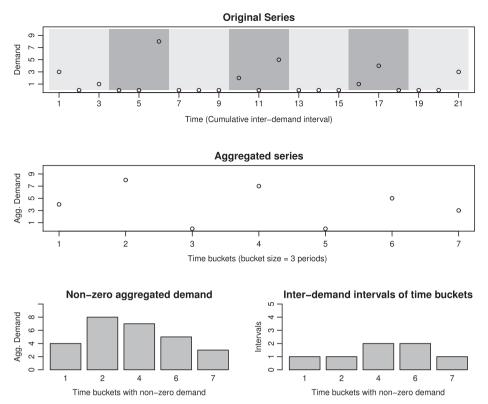
Fig. 2. Illustrative example of the non-overlapping temporal aggregation for an intermittent demand series. The aggregation level (time bucket size) is equal to three periods.

Table 1
Croston's decomposition of the intermittent demand series in Fig. 1.

(1) Period	(2) Non-zero demands	(3) Inter-demand intervals	(4) Cumulative non-zero demands
1	3	1	3
3	1	2	4
6	8	3	12
10	2	4	14
12	5	2	19
16	1	4	20
17	4	1	24
21	3	4	27

inverse series gives the estimate of the periods that will pass before a pre-fixed quantity of units is demanded. As a result this alternative view for intermittent demand time series may be very useful when dealing with "clump" sizes (Ritchie and Kingsman, 1985; Vereecke and Verstraeten, 1994).

3.2. Inverse ADIDA

An important property of the inverse intermittent demand series is that, when focusing to Croston's method estimates (as calculated by Eq. (4)), the inverse of its forecasts is equal to the forecasts of the original series. However, the application of the ADIDA framework on the inverted series will end up with different estimates compared to the application of ADIDA on the original series. While the direct application of ADIDA on the original data creates fixed time buckets with variable aggregated demand, the non-overlapping temporal aggregation on the inverted series calculates the aggregated sizes of the inter-demand intervals over fixed 'demand buckets'.

In essence, by considering the inverse series, we are now able to reduce the observed variance in the demand sizes and focus on

the estimation of the inter-demand intervals. We call the application of the ADIDA framework on the inverted series *Inverse ADIDA* (iADIDA). Fig. 4 illustrates the iADIDA framework, using an aggregation level that corresponds to demand buckets of size 3. The inverted series and its aggregated counterpart are presented in the two first panels. Once the aggregation of the demand buckets has been performed, we can consider the inverse of the aggregated inverted series (third panel). Both the aggregated inverted series and its inverse can be decomposed in the non-zero demand buckets and the inter-demand-buckets aggregated intervals as depicted in the two bottom bar-graphs of Fig. 4.

With regard to the selection of the aggregation level, in both ADIDA and iADIDA, any value greater than unity will reduce the variance in the sub-series of the inter-demand intervals and the demand sizes respectively. In the case of ADIDA, a data-driven aggregation level could match the mean inter-demand interval, so the intermittence will be (almost) removed (Petropoulos and Kourentzes, 2015). A data-driven aggregation level in the case of iADIDA could match the mean demand size per series. Also, and as already mentioned, Nikolopoulos et al. (2011) provide a managerially driven selection for the aggregation level in ADIDA framework, so that this matches the lead-time plus the review period, as cumulative forecasts over that time horizon are required for stock control purposes. The managerially driven equivalent in iADIDA would be the alignment of the aggregation level with the "clump size" in which SKUs are distributed.

In the rest of this paper we test the performance of the proposed iADIDA method and discuss its value from a managerial perspective.

4. Experimental design

We consider two real-life data sets that have been used widely in previous research for intermittent demand. The first consists of

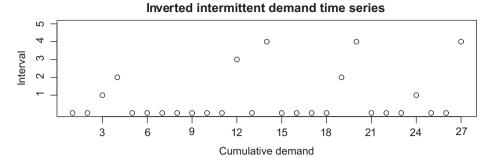**Inverted intermittent demand time series**

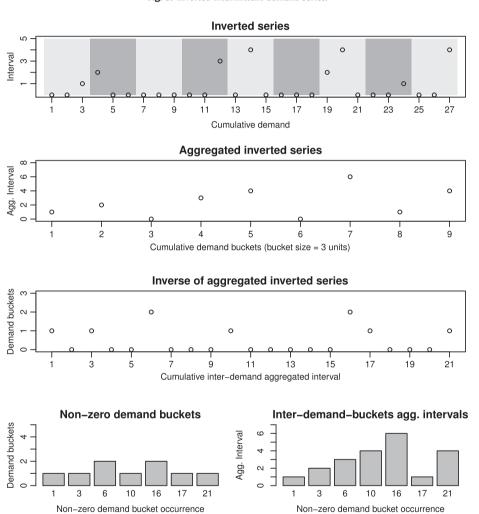Fig. 3. Inverted intermittent demand series.

Fig. 4. An illustrated view of Inverse ADIDA framework. The aggregation level (demand bucket size) is equal to three units of demand.

the demand histories of 3000 SKUs from the automotive industry over two years (Syntetos and Boylan, 2005; Syntetos et al., 2005), while the second refers to the demand of 5000 SKUs from the Royal Air Force (RAF) over a period of 7 years (Teunter and Duncan, 2009; Syntetos et al., 2009). Demand is recorded in monthly frequency. The descriptive statistics for both data sets are provided in Tables 2 and 3 respectively.

We employ a rolling origin evaluation scheme. Holding out the n last data points, we produce the one-step-ahead forecast (forecast horizon equals to one period). Then, one additional data point is added into the in-sample and a new forecast is calculated from the next origin. This procedure is repeated until the produced forecast corresponds to the last available observation.

The n is set equal to 12 periods for the RAF data set. In other words, the first origin of the rolling evaluation is the 72nd period. Previous research has considered a similar evaluation window on the same data set (Petropoulos and Kourentzes, 2015). However, due to limited sample size, n is set equal to 6 periods for the automotive data set (the first origin of the rolling evaluation is the 18th period).

Four different error measures are used to compare the forecasting performance of the different approaches:

- Scaled Mean Error (sME), which refers to the mean signed error scaled by the mean demand of each series. This error measure is suitable for measuring the bias of the different methods.

Table 2
Descriptive statistics of the of the automotive data set.

3000 SKUs	Demand sizes(units)		Inter-demand intervals (months)		Demand per period (units/month)	
	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.
Min	1.00	0.00	1.04	0.21	0.54	0.50
25%ile	2.05	1.14	1.10	0.30	1.46	1.32
Median	2.89	1.76	1.26	0.52	2.33	1.92
75%ile	5.00	3.36	1.41	0.73	4.17	3.50
Max	193.75	101.41	2.00	1.59	129.17	122.75

Table 3
Descriptive statistics of the of the RAF data set.

5000 SKUs	Demand sizes(units)		Inter-demand intervals (months)		Demand per period (units/month)	
	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.
Min	1.00	0.00	3.82	0.00	0.04	0.19
25%ile	1.56	0.81	7.27	5.43	0.15	0.54
Median	3.83	3.06	9.00	6.93	0.37	1.45
75%ile	11.33	9.31	11.57	8.63	1.15	4.43
Max	668.00	874.42	24.00	16.46	65.08	275.71

- Scaled Mean Absolute Error (sMAE), which refers to the mean unsigned error scaled by the mean demand of each series. This error is appropriate for evaluating the accuracy of the estimates.
- Scaled Mean Squared Error (sMSE), which is the scaled equivalent of Mean Squared Error and is calculated by dividing each squared error by the squared average of the actual demands.
- Mean Absolute Scaled Error (MASE), where the mean absolute error is scaled by the in-sample performance of random-walk (Hyndman and Koehler, 2006). Like sMAE, this measure is an appropriate indicator of the methods' accuracy.

For more details on the selected error measures, the reader is encouraged to refer to Petropoulos and Kourentzes (2015). The error measures employed are meant to demonstrate the relative performance of the proposed iADIDA method in terms of three different dimensions: bias (sME), accuracy (sMAE and MASE) and variance (sMSE).

Apart from the evaluation with different error metrics, an inventory simulation was performed to assess the alternative approaches in terms of stock control performance. Following the inventory simulation design by Kourentzes (2014), we simulate intermittent demand series of 1180 periods. The series are simulated assuming Bernoulli demand arrivals and negative binomial demand sizes (Petropoulos et al., 2014), using the values of the inter-demand interval and the demand squared coefficient of variation from real time series of the two data sets considered. The first 60 observations are used as the training set, while the next 1000 are treated as the burn-in period. The last 120 observations are considered as the test-set, where the inventory performance is measured.

Following the recommendations of Teunter and Sani (2009), we use an order-up-to-policy that is common in practice. We set four different target service levels (80%, 90%, 95% and 99%) to measure the performance at various reasonable settings. The inventory review period is set to 1 period (month). To initialise the inventory simulation we assume the initial stock to be equal to the order-up-to level and zero orders in the system. Since these initial values are ad hoc we use a long burn-in period in our simulation to eliminate any artefacts in the results due to the choice of initial values. We consider lead times of 1 and 5 periods and to simplify the

simulation it is assumed that any out-of-stocks do not produce additional demand in the following periods and are serviced by competitors. We keep track of two quantities, the holding stock, i.e. how much stock was stored during the simulation evaluation period, and the backlog, which is the sum of unserviced units within the evaluated simulation period. Both quantities are divided by the mean of the actual demand to make them scale independent and allow for summarisation of the results across time series.

5. Evaluation

The two selected data sets provide an ideal platform for evaluating iADIDA as they pose different types of demand. Tables 2 and 3 show that the automotive data set demonstrates lower demand intervals than the RAF data set, so the latter is far more intermittent than the former. At the same time, the demand-per-period inter-quartile range for the automotive data set is 271% higher than the respective figure of the RAF data set. Thus, the demand of the automotive data set could be characterised as more erratic. Therefore, we expect iADIDA to perform better in the automotive data set rather than the RAF one.

Tables 4 and 5 present the results on the empirical evaluation for the Automotive and RAF data sets respectively. The performance of iADIDA is contrasted to that of ADIDA for various aggregation levels. In total, we consider three variations of iADIDA. The first one, iADIDA(2), utilises a fixed aggregation level of 2 demand units. The second one, iADIDA(mean), refers to the data-driven aggregation level discussed in Section 3, where the aggregation level of each series is set equal to the mean demand size. The third one, iADIDA(max), sets the aggregation level equal to the maximum demand size observed in each series. In the latter case, the 'intermittence' of the demands in the inverse series is completely removed. In all cases of ADIDA and iADIDA, the underlying forecasting method is suggested by the PK classification scheme (selecting between SES, Croston's method or SBA).

Also, three simple methods (Naive, Croston and SBA) and the PK classification scheme applied on the original data are used as benchmarks. In all cases apart from the Naive method, the performance of each method is provided both for fixed and optimised values of the smoothing parameters. The results for the methods using optimised smoothing parameters are presented in brackets. The fixed value of both smoothing parameters (α_z and α_p) is set to 0.1.

The performance of both Croston's method and SBA is significantly better than that of the Naive, with the latter producing negatively biased forecasts (note that $e = y - \hat{y}$). However, the Naive method achieves superior accuracy in the RAF data set. This is due to the fact that the series in these data sets are highly intermittent (the median inter-demand interval is 9), meaning that

Table 4
Empirical performance results for the different approaches considered (automotive data set).

Method	sME	sMAE	sMSE	MASE
Naive	0.031	0.942	2.621	1.047
Croston	−0.045 (−0.023)	0.780 (0.771)	1.969 (**1.949**)	0.869 (0.858)
SBA	0.006 (−0.017)	0.769 (0.770)	1.966 (1.950)	0.855 (0.857)
PK	−0.004 (−0.019)	0.772 (0.771)	1.967 (1.952)	0.859 (0.858)
ADIDA(2)	−0.013 (−**0.010**)	0.776 (0.768)	1.973 (1.954)	0.863 (0.855)
ADIDA(6)	−0.023 (−0.043)	0.772 (0.791)	**1.952** (1.993)	0.859 (0.881)
iADIDA(2)	**0.003** (−0.015)	0.771 (0.772)	1.970 (1.956)	0.858 (0.859)
iADIDA(mean)	0.012 (0.218)	0.770 (**0.747**)	1.973 (2.004)	0.856 (**0.830**)
iADIDA(max)	0.012 (−0.013)	**0.764** (0.771)	**1.952** (1.961)	**0.851** (0.858)

Table 5
Empirical performance results for the different approaches considered (RAF data set).

Method	sME	sMAE	sMSE	MASE
Naive	0.263	**1.662**	92.830	**0.931**
Croston	−0.123 (−0.024)	1.961 (1.877)	81.192 (81.099)	1.050 (1.005)
SBA	−0.067 (−0.016)	1.913 (1.871)	81.172 (81.103)	1.025 (1.001)
PK	−0.067 (−0.016)	1.913 (1.871)	81.172 (81.103)	1.025 (1.001)
ADIDA(2)	−0.061 (−0.006)	1.908 (1.862)	81.182 (81.102)	1.021 (0.996)
ADIDA(12)	**−0.013** (0.044)	1.869 (1.822)	81.234 (81.160)	0.999 (0.974)
iADIDA(2)	−0.055 (−**0.001**)	1.903 (1.856)	81.156 (81.139)	1.019 (0.994)
iADIDA(mean)	−0.050 (0.130)	1.899 (1.747)	81.129 (**81.092**)	1.016 (0.935)
iADIDA(max)	−0.049 (−0.018)	1.897 (1.872)	**81.067** (81.231)	1.015 (1.002)

it conveniently produces spot-on zero forecasts which results in under-stocking (highly positive sME).

In both data sets, SBA results in overall better performance compared to Croston's method. This was expected, as SBA produces less biased forecasts. The PK classification scheme offers a balanced performance in the automotive data set, while the high values in the average inter-demand interval of the RAF data set render the classification scheme identical to SBA. As expected, the improvements offered by the ADIDA framework are more apparent in the RAF data set, where the variance of the inter-demand intervals is larger compared to the automotive one. If the zero forecasts of Naive method are disregarded, ADIDA(12) clearly provides the best performance in terms of accuracy for the RAF data set. Confirming the results by Kourentzes (2014), the use of optimised smoothing parameters compared to ad hoc ones leads to improved performance in the case of RAF data set.

All variants of iADIDA show very promising performance which is comparable to the rest of the benchmarks considered across all considered metrics and both data sets. As it was expected, the iADIDA framework performs especially well in the case of the automotive data. iADIDA(mean) and iADIDA(max) yield the most accurate results for optimised and fixed smoothing parameters respectively.

It is worth noticing that in both data sets iADIDA results in lower sMSE values and as such the resulted forecasts have less variance. From a managerial and decision making perspective, this is an extremely important property as it allows for a smoother ordering and decision making process (if based and driven on the derived forecasts). This is also demonstrated to some extent in the second part of the analysis where we evaluate the stock control performance of the different approaches based on simulated data. The efficiency curves (Figs. 5 and 6) show that iADIDA performs by and large at the same levels as the other methods on stock control performance.
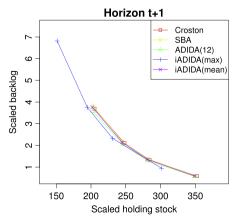
Lastly, note that our results are not always aligned with our theoretical expectations. iADIDA(mean) seems to suffer more than the other models in the bias front when the smoothing parameters are optimised. This latter finding may be attributed to over-smoothing. The 'mean' version of iADIDA method applies an aggregation level that is equal to the mean demand size of each series. So, iADIDA(mean) performs a first layer of smoothing and is tuned to work on 'average' conditions. Superimposing a layer of optimisation could result in solutions that lie on the over-optimisation spectrum. This, however, is not the case for iADIDA (2) and iADIDA(max) that do not apply that first layer of smoothing and this empirical finding provides some more validity to the aforementioned argumentation. The result, however, remains equally strong: there is always an iADIDA model that performs on par or better than well-researched benchmarks and as such the way forward should be a selection procedure that includes a pool of iADIDA variants.

With regard to implications to practice one important question remains: when does the method work? iADIDA performs in a similar way to the original ADIDA, in the sense that it helps methods to improve their performance. It is a *method self-improving-performance mechanism*. In this specific instance iADIDA improves the performance of PK. PK performs a selection between Croston's method, SBA and SES based on the values of the mean inter-demand interval and the squared coefficient of variation. So, iADIDA is expected to work under the same conditions that the underlying extrapolation methods work and moreover offer further gains.

We stress that the new proposition is neither a panacea for time series forecasting nor a way to improve the performance of any method under any conditions. There is a condition that favours the application of the iADIDA framework and thus the performance boosting is amplified. This is when a time series has high data-volume variance. We empirically observed such performance boosting in the automotive data set that demonstrates such time series characteristics. In any case, further investigation is needed for a broader context of selective applicability, a question that is still open for many of the intermittent demand methods.

6. Concluding remarks

In this paper we proposed an innovative way of temporally aggregating intermittent demand data in order to reduce the variance of the demand. The new approach (iADIDA) is based on the inversion of the intermittent demand series, which is feasible through Croston's decomposition. In essence, iADIDA is expected to work better when a time-series presents high data-volume variance as methodologically it directly results in reduction of that
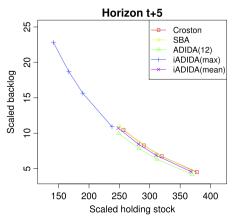

Fig. 5. Inventory scaled trade-off curves for the simulated series based on the automotive data set.
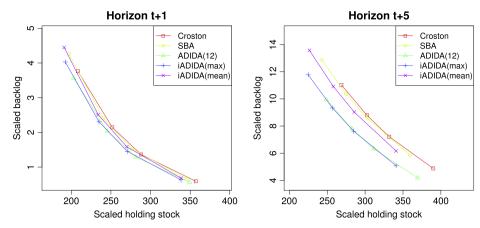
Fig. 6. Inventory scaled trade-off curves for the simulated series based on the RAF data set.

specific type of variance, in contrast to the ADIDA framework (Nikolopoulos et al., 2011) that results in reduction of the inter-demand interval variance. We expect that this new approach will result in improvements related to reduction of the forecasts' variance and possibly increased accuracy performance primarily in the case of 'erratic' demand and less for 'lumpy' demand, as defined by Syntetos et al. (2005).

We argue that it should not be universally applied in a data set, but selectively for the subset containing the time series with the highest data-volume variance. This could be imposed via an ad hoc threshold, such as the data-volume variance third quartile, or a more informed and empirically driven criterion through a competition in a holdout. We intend to investigate this question further in future research. It is important to note that this is inexorably linked with the selection of the basic intermittent demand forecasting method selection that remains an open research question.

Given that the proposed algorithm is expected to perform better under a selectivity criterion, the question that remains is what would be the nominally selected forecasting approach for the rest of the data that present average or low data-volume variance. This could be resolved through a selection protocol along the lines of the 'horses for courses' approach as discussed by Petropoulos et al. (2014) where SBA and TSB would be the natural dominant choices.

In this work we focused on the performance of the new approach using real data and a variety of error metrics, as well as simulated data and inventory stock control evaluation. The theoretically expected impact on the forecasting performance for smooth and intermittent series remains out of the scope of this study and thus we leave it for potential further research. Nonetheless, this work demonstrates that there is merit in re-interpreting the idea of Croston's decomposition. The literature has so far focused on interpreting intermittent time series is a single way that has persisted for more than four decades. In this work we argue that the proposed inverse interpretation is useful and allows us to handle the information contained in the time series in novel ways, leading to gains in forecast performance. We anticipate that this will spur new research in modelling intermittent time series.

References

Croston, J.D., 1972. Forecasting and stock control for intermittent demands. Oper. Res. Q. (1970–1977) 23 (3), 289–303.

Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. Int. J. Forecast. 22 (4), 679–688.

Kostenko, A.V., Hyndman, R.J., 2006. A note on the categorization of demand patterns. J. Oper. Res. Soc. 57, 1256–1257.

Kourentzes, N., 2013. Intermittent demand forecasts with neural networks. Int. J. Prod. Econ. 143, 198–206.

Kourentzes, N., 2014. On intermittent demand model optimisation and selection. Int. J. Prod. Econ. 156, 180–190.

Kourentzes, N., Petropoulos, F., 2016. Forecasting with multivariate temporal aggregation: the case of promotional modelling. Int. J. Prod. Econ. 181 (PA), 145–153. http://dx.doi.org/10.1016/j.ijpe.2015.09.011.

Kourentzes, N., Petropoulos, F., Trapero, J.R., 2014. Improving forecasting by estimating time series structural components across multiple frequencies. Int. J. Forecast. 30 (2), 291–302.

Nikolopoulos, K., Syntetos, A.A., Boylan, J.E., Petropoulos, F., Assimakopoulos, V., 2011. An aggregate–disaggregate intermittent demand approach (ADIDA) to forecasting: an empirical proposition and analysis. J. Oper. Res. Soc. 62 (3), 544–554.

Petropoulos, F., Kourentzes, N., 2015. Forecast combinations for intermittent demand. J. Oper. Res. Soc. 66, 914–924.

Petropoulos, F., Makridakis, S., Assimakopoulos, V., Nikolopoulos, K., 2014. Horses for courses in demand forecasting. Eur. J. Oper. Res. 237, 152–163.

Petropoulos, F., Nikolopoulos, K., Spithourakis, G., Assimakopoulos, V., 2013. Empirical heuristics for improving intermittent demand forecasting. Ind. Manag. Data Syst. 113 (5), 683–696.

Ritchie, E., Kingsman, B., 1985. Setting stock levels for wholesaling: performance measures and conflict of objectives between supplier and stockist. Eur. J. Oper. Res. 20, 17–24.

Rostami-Tabar, B., Babai, M., Syntetos, A., Ducq, Y., 2013. Demand forecasting by temporal aggregation. Naval Res. Logist. 60 (6), 479–498.

Rostami-Tabar, B., Babai, M., Syntetos, A., Ducq, Y., 2014. A note on the forecast performance of temporal aggregation. Naval Res. Logist. 61, 489–500.

Spithourakis, G., Petropoulos, F., Babai, M.Z., Nikolopoulos, K., Assimakopoulos, V., 2011. Improving the performance of popular supply chain forecasting techniques. Supply Chain Forum, An International Journal 12 (4), 16–25.

Syntetos, A., Babai, M., Dallery, Y., Teunter, R., 2009. Periodic control of intermittent demand items: theory and empirical analysis. J. Oper. Res. Soc. 60, 611–618.

Syntetos, A.A., Boylan, J.E., 2001. On the bias of intermittent demand estimates. Int. J. Prod. Econ. 71 (May (1–3)), 457–466.

Syntetos, A.A., Boylan, J.E., 2005. The accuracy of intermittent demand estimates. Int. J. Forecast. 21 (2), 303–314.

Syntetos, A.A., Boylan, J.E., Croston, J.D., 2005. On the categorization of demand patterns. J. Oper. Res. Soc. 56 (5), 495–503.

Teunter, R.H., Duncan, L., 2009. Forecasting intermittent demand: a comparative study. J. Oper. Res. Soc. 60, 321–329.

Teunter, R.H., Sani, B., 2009. On the bias of Croston's forecasting method. Eur. J. Oper. Res. 194 (1), 177–183.

Teunter, R.H., Syntetos, A.A., Babai, M.Z., 2011. Intermittent demand: linking forecasting to inventory obsolescence. Eur. J. Oper. Res. 214 (3), 606–615.

Vereecke, A., Verstraeten, P., 1994. An inventory management model for an inventory consisting of lumpy items, slow movers and fast movers. Int. J. Prod. Econ. 35, 379–389.

Willemain, T.R., Smart, C.N., Shockor, J.H., DeSautels, P.A., 1994. Forecasting intermittent demand in manufacturing: a comparative evaluation of Croston's method. Int. J. Forecast. 10 (4), 529–538.