Accepted 3 January 2020 for publication in European Journal of Operational Research

We need to talk about intermittent demand forecasting[1,2]

Konstantinos Nikolopoulos

forLAB www.forlab.eu
Professor and Chair in Business Analytics
Bangor Business School
Prifysgol Bangor University
Hen Colleg 2.06
Bangor, Gwynedd
Wales, UK
kostas@bangor.ac.uk
+44 1248 383796

[1] The title of the article is inspired by Scriver's book and the respective film: "We need to talk about Kevin' and analogies to the main theme and character of the story, Kevin: people often miss or ignore the 'obvious'. This I s an argument made very emphatically in this paper on the lack of practitioners and academic to use and see the potential of intermittent demand forecasting methods outside the strict application area of spare parts and inventory management.

2 Earlier versions of this work have presented and appeared in (chronological order):

a) Nikolopoulos, K. & Syntetos, A.A. (2013), "Forecasting Black, Gray & White Swans…", 2013 DSI Annual Meeting, November 16-19, 2013, Baltimore, Maryland, USA,

b) Nikolopoulos, K., Syntetos, A.A., Batiz-Lazo, B., (2013), "Forecasting Black (& White) Swans…", OR55, 3–5 September 2013. The University of Exeter, Exeter, England, United Kingdom. Keynote Presentation for the 'Forecasting stream', and

c) K. Nikolopoulos & F. Petropoulos (2015), "Forecasting, Foresight and Strategic Planning for Black Swans", Bangor Business School Working Paper Series, BBSWP/15/03. https://www.bangor.ac.uk/business/research/documents/BBSWP-15-03.pdf

I thank Professor Fotios Petropoulos from University of Bath School of Management and Professor Aris A. Syntetos from Cardiff University Business School for their views and reviews on earlier versions of this paper. I thanks Dr Fotios Petropoulos for preparing some of the simulations and empirical evaluations in order to support my thoughts in this article. These ideas have also been presented and influenced by the audiences in a series of invited talks – in the form of free thoughts under the title "We need to talk about Kevin", more recently in Leeds University Business School (2018), Cardiff Business School (2018) and the work foundation in London (2018).

Accepted 3 January 2020 for publication in European Journal of Operational Research

We need to talk about intermittent demand forecasting

Abstract

Operational Research (OR) is the 'science of better'. People constantly try to get better, in practically all aspects of their personal and professional life, and thus OR is de facto a ubiquitous science. What might however not be that clear, is that the way we improve is driven by the very OR science, and the scientific results that constitute the respective body of literature, theory and practice. Of all the tasks, that can be related to the OR science, the one that more frequently we do, is forecasting. We do constantly try to estimate what is coming next, and we drive our decisions for each and every situation based on these forecasts: from where to put our money to who will be the best surgeon to operate us. Within the broad boundaries of OR, forecasting stands out as the most ubiquitous sub-discipline. In the forecasting literature, a lot of attention has been given to modelling fast-moving time series and building causal models; however, very limited attention has been given to intermittent time series and intermittent demand forecasting. In this research, we advocate for the broader use of intermittent demand forecasting methods for forecasting special events, as a simpler, faster, and robust alternative to more complex non-OR models. Furthermore, in a foresight context, we argue for a novel way of deciding the strategic planning horizon for phenomena prone to appearance of special events.

Keywords: Forecasting; Ubiquitousness; Intermittent Demand; Special Events; Foresight;

1. Introduction: on the ubiquitousness of Operational Research.

Operational Research (OR) is the 'science of better'[3]. Everybody, constantly, tries to get better, in everything, and thus OR is de facto a ubiquitous science. Our daily lives are positively influenced by OR; the credit however does not necessarily go to the discipline. What might not be clear, is that the way we improve, is driven by the OR science and the scientific results that constitute the respective body of literature, theory and practice. In academia and practice, we need advocate more strongly for this everyday successful application of OR in our lives.

The lack of acknowledgment of the contribution of OR should not come as a surprise. Even within academia, we do not even agree how the discipline should be called: operational research, operations research or management science. Thus it makes perfect sense that we cannot agree what it entails, and what kind of practical problems it is meant to be solving? Moreover, the birth of analytics in 2007 (Davenport and Harris) spayed even more gasoline on the fire: an amalgamation of mathematics, management, and computer science storm out industry and academia.

Although proof by example is a common fallacy, we have no other option rather than throwing as many examples as possible of how OR changes our lives, and therefore provide more evidence of the ubiquitous nature and use of OR.

[3] http://www.scienceofbetter.co.uk/

Within the ubiquitous OR discipline, *forecasting* is the finest example (Perera et al., 2019; Fildes et al., 2008; Fildes and Nikolopoulos, 2006). Applications span from supply chain forecasting (Rostami-Tabar et al., 2019; Syntetos et al. 2016), inventory management (Babai et al. 2012), marketing (Nikolopoulos et al. 2007), deep learning (Punia et al. 2019) and big data (Nikolopoulos and Petropoulos, 2018), tourism (Petropoulos et al., 2006) , healthcare (Athanasopoulos et al., 2017; Nikolopoulos et al. 2015), to human judgment and behavioral operations (Petropoulos et al., 2018; Kremer et al., 2016) and super humans (Katsagounos et al., 2020). The later referring to the famous work on superforecasting of Tetlock, Mellers (Tetlock and Gardner, 2015). For every situation and context, there is a forecasting model to drive informed decisions: for every course, there is a horse… (Petropoulos et al. 2014)

At the individual level, of all the tasks that can be associated with the OR science, the one task that we more frequently do, is forecasting. We try to estimate what is coming next, and we drive our decisions based on these forecasts: from where to put our money to who will be the best surgeon to operate us.

The research team in Uber, that recently triumphed among more than sixty academics and professional participants in the latest forecasting competition - the M4 competition where the task was to forecast blindly and unsupervised 100,000 times series (Makridakis et al. 2018), are elaborating publicly[4] on this line of argument:

[4] https://eng.uber.com/forecasting-introduction/

> *"Forecasting is ubiquitous. In addition to strategic forecasts, such as those predicting revenue, production, and spending, organizations across industries need accurate short-term, tactical forecasts, such as the amount of goods to be ordered and number of employees needed, to keep pace with their growth"*

2. Literature review: on the importance of intermittent demand forecasting methods.

The title of this article is inspired by Scriver's popular fiction book: "We need to talk about Kevin". An analogy is made in this article to the main character of the story, 'Kevin': people often miss the forest for the trees, and in the case of Kevin missed all the signs that built up to create a mass killer. This is an argument made very emphatically in this paper, about the lack of practitioners and academics to use and see the potential of intermittent demand forecasting methods outside the strict application area of spare parts and inventory management. So in this section we strongly argue and bring in respective literature that suggests that we need to talk about intermittent demand forecasting methods.

In the forecasting literature a lot of attention has been given to modelling fast-moving time series (Nikolopoulos and Thomakos 2019; Petropoulos et al., 2014) as well as using causal models when cues of information are available (Bozos and Nikolopoulos, 2011; Nikolopoulos et al. 2007). However, limited attention has been given intermittent demand series[5] and respective forecasting methods. For this aforementioned lack of academic attention, we can blame the widespread perception that intermittent demand forecasting methods are only useful for spare parts demand forecasting. Even if that was true, this is still a huge omission as 60% of any inventory consists of spare part where intermittent demand is pertinent (Syntetos et al. 2016), and these can be very expensive parts as many papers deal with RAF spare parts (for example Nikolopoulos et al., 2011).

Another reason to explain the lack of academic research, is that it is twice as difficult as any other forecasting problem; thus academics tend to avoid so difficult problems in the early stages of their career. In intermittence, we have two sources of uncertainty: the sporadic nature of the actual demand volume, plus the timing of demand arrivals. Furthermore, to make things even more difficult, the assumption is that intermittent data are i.i.d. so there should be no time series characteristics evident there (Petropoulos et al., 2014). There is research arguing against that as Altay et. Al. (2018, 2012) identified trends and seasonality in intermittent series, while Nikolopoulos et al. (2016) discussed the existence and identification of local patterns driven by buyers and suppliers behaviors.

[5] Referred also in the literature as: *sporadic demand* and *count series*.

There have been very few forecasting methods developed specifically for intermittent data over the last five decades (Syntetos et al., 2006; Petropoulos et al., 2014). The first and defining proposition was from Croston (1972), followed many years later by Syntetos and Boylan (SBA, 2001) and Willemain et al. (2004); and more recently Teunter et al. (TSB, 2011) and Nikolopoulos et al. (ADIDA, 2011). The latter being more a method-improving non-overlapping temporal aggregation framework rather than a new method per se.

This latter idea was further pursued by many researchers including Kourentzes et al. (MAPA, 2014) and Petropoulos et al. (iADIDA, 2016). Syntetos et al. (2015) and Petropoulos et al. (2014) provided extensive empirical evaluations and simulation of methods for intermittent demand. While Syntetos et al. provided a standard forecasting competition, Petropoulos et al. went one-step further to identify what time series characteristics favor every method.

In all proposed application so far in the literature, intermittent demand estimators were designed and tested to forecast spare part demand and respectively drive stock control decisions. There has not been consideration, to the best of my knowledge, on how these methods can be used out of the core context of OR and inventory management.

The line of argument we follow in this research is that intermittent demand forecasting methods can help up us deal with the worst form of uncertainty: rare special events (Fildes and Nikolopoulos, 2006) like black and gray swans (Taleb, 2007). We adopt Taleb's (2007) classification of: a) white swans, the mainstream of a time series that are usually predictable with time series and/or causal methods, b) gray swans, that are rare but expected and "*somewhat predictable, particularly to those who are prepared for them and have the tools to understand them*", and c) black swans, the truly unexpected special events.

We particularly focus on the latter two classes and:

a) advocate for the use of Operational Research (OR) forecasting tools i.e. intermittent demand forecasting methods for forecasting special events black and gray swans, as a simple, fast and robust alternative econometric probabilistic methods like Extreme Value Theory (EVT) (Leadbetter, 1991),

b) for the sake of the argument and illustration we demonstrate the use of a rather popular forecasting paradigm: the Naive method through the ADIDA non-overlapping temporal aggregation method-improving framework (Nikolopoulos et al., 2011) – however any other popular method could be used like SBA or TSB, and

c) propose a new way for deciding the strategic planning horizon for phenomena prone to the appearance of special events like black and gray swans.

3. Forecasting special events with intermittent demand forecasting methods.

Intermittent demand patterns are evident in real life and have been studied for thousands of years: 2500 years ago in Ancient Greece, Archimedes described the intermittent nature of earthquake occurrences, and he was not the first one to use such modeling in the past. Since the start of this millennium and the article from Syntetos and Boylan (2001), there has been a significant increase in the interest in intermittent demand and respective classification, modelling and forecasting methods in this context; to attest to that, the seminal work from Croston (1972) has received in this recent period more than 150 citations.

We argue that this interest could be further augmented, via a time series decomposition lens in a similar fashion like Leadbetter (1991) through isolating 'Peaks over Threshold' data points. This would lead to a new intermittent demand series, that has identical characteristics with the series that Croston aspired to mathematically model and forecast back in 1972. In essence one could use intermittent demand modeling and forecasting methods in order to research problems in finance, business, economics and social sciences: in fact in any other discipline. The focus will be given in the peaks-over-threshold[6] of this series, that are the points that carry most of the risk and are of eminent interest to academics and practitioners.

[6] Peaks-Over-Threshold and Points-Over-Threshold are terms that are used interchangeable in the literature and practice, with the former more common in academia.

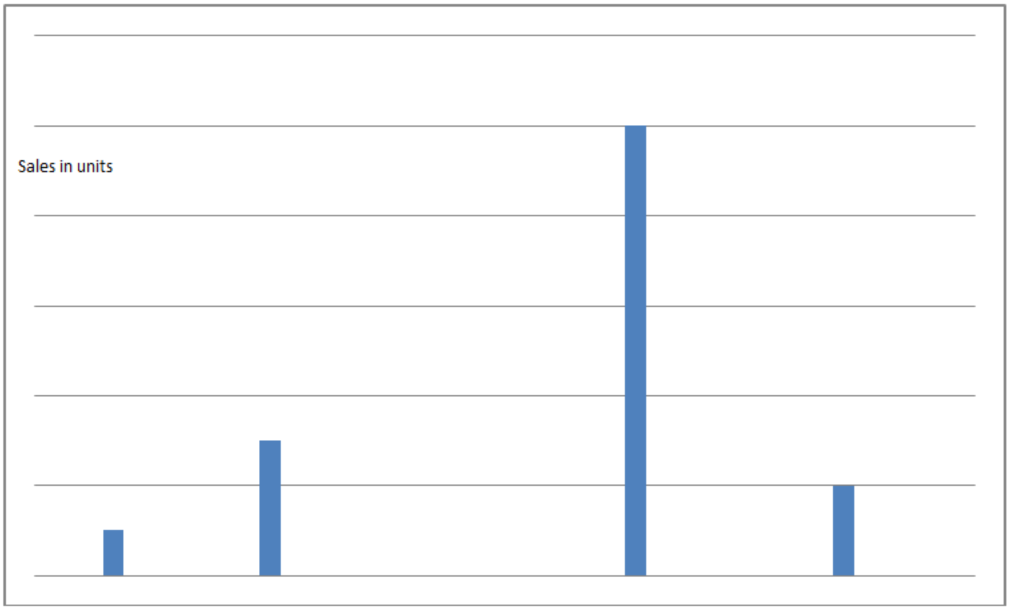


Figure 1. Illustration of a time series of intermittent demand.

An intermittent demand series is illustrated in figure 1, characterized by periods with demand interchanging with periods of no demand at all. Demand volume (when realized) comes with significant variation. Modelling such a series is very challenging, as there are two things to forecast:

- When the next demand period is going to be realized?
- What will be the volume of this demand?

3.1 A decomposition lens: peaks-over-threshold series.

In this section we argue through a series of illustrative examples, how any times series could be decomposed into two subseries: one containing the baseline (*white swans*) and one containing the special events, the extremes (*gray* and *black swans*). For the former traditional forecasting approaches could be used (Petropoulos et al. 2014, Nikolopoulos and Thomakos, 2019). For the latter, we propose the use of intermittent demand methods. It is worth mentioning, that these extreme data points capture most of the risk associated with a phenomenon, and thus it is very important to be able to effectively model and forecast them.

In figure 2 we provide three illustrative examples of creating peaks-over-threshold intermittent demand series from data for global earthquakes, the S&P 500 stock market index and influenza cases. The threshold inevitably varies in between application areas: what constitutes an epidemic over a normal flu season is a different threshold from what a high-risk period in a volatility series is or what is considered a major earthquake.

The threshold can be set through fitting a distribution to the data (Pareto distribution, Generalized Pareto Distribution (GPD) or even normal distribution) or by simply applying a rule of thumb (Pareto 90/10 or 99/1). Soleari et al. (2017) have also proposed a methodology for automatic threshold estimation using goodness of fit p-value.

Another popular approach is through an a-priori defined number of standard deviations around the mean; with common choices being 3 to 5 standard deviations over the mean. For the normal distribution if the threshold is set to 3σ then it would identify as extremes only 0.01% of the data. More often practitioners set these thresholds.

Finally, a common also approach does not require a-priory ad-hoc threshold is through exploratory data analysis; in that approach a boxplot with an outer fence is constructed, and any data points outside the outer fence would be considered special events, and as such part of the separate intermitted peaks-over–threshold series. The advantage here is that the mainstream 50% of the data within the two quartiles (Q1 and Q3) would define the threshold(s):

- threshold for high values: > Q3 + 3 (Q3-Q1) , and
- threshold for low values: < Q1 – 3 (Q3-Q1) .

In any case, the extreme data are usually handled separately (or just eliminated) from the rest of the mainstream time series data. It needs to be emphasized that how the threshold is decided this is not one of the main objectives over the article. The threshold can be set with many different ways: and then an intermittent series will be created, with more or less intermittence[7]. We will then need to forecast this newly derived time series. We therefore propose and argue in this article for the use intermittent demand forecasting methods as a viable and promising alternative for the subject matter.

---

[7] Intermittence is usually measure through the ratio of the number of points where demand is realized (referred in the literature as 'issue points') over the number of points where there is no demand at all. The less the ration, the more intermittent the time series is.

In our three examples:

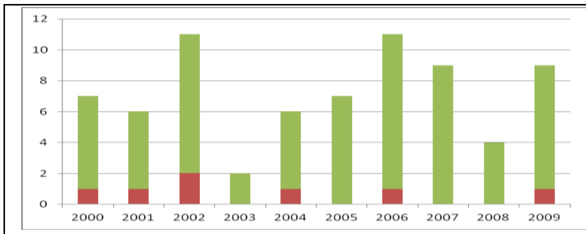
a) for the earthquake historical data the threshold is set at 7 points of the Richter scale, as any earthquake over that usually is quite catastrophic, with significant casualties and catastrophic economic consequences (and usually is followed by a Tsunami warning in the recent years),

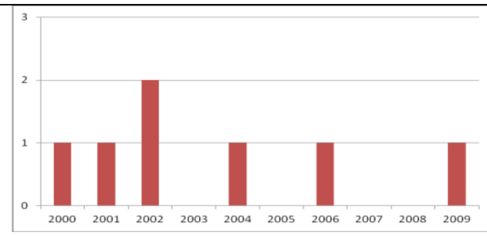
b) for the volatility series where there are no specific units of measurement, the threshold is decided following a visual inspection, and in this instance it was set to 0.02,

c) in the case of influenza thresholds vary and many rely on qualitative criteria of what is considered a pandemic[8]; in the illustrative example if more than 5% visits in primary care units and GPS are for ILIs (Influenza Like Illnesses) then that would capture two recent special events. If the threshold set to 7% then it would identify only the 2009 H1N1 Swine flu as a special event – a pandemic in this case.

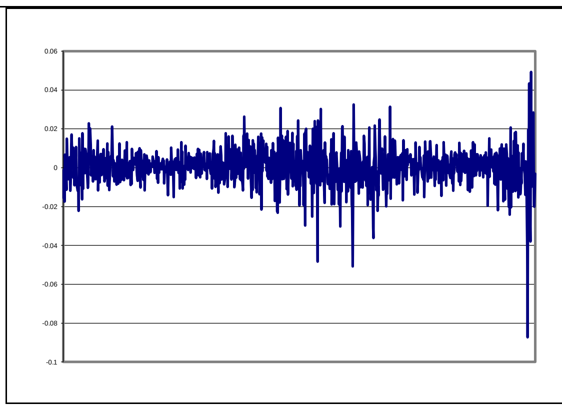
The intermittent series in the right column of 2a, 2b and 2c in figure 2 have similar characteristics, fully consistent with what is traditional perceived and classified as intermittent demand (figure 1), and thus more or less advanced intermitted demand forecasting methods could be employed (Petropoulos and Kourentzes, 2014; Nikolopoulos et al., 2011; Willemain et al., 2004; Croston, 1972).

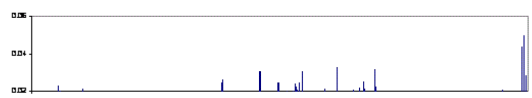
[8] https://www.who.int/bulletin/volumes/96/8/18-211508/en/



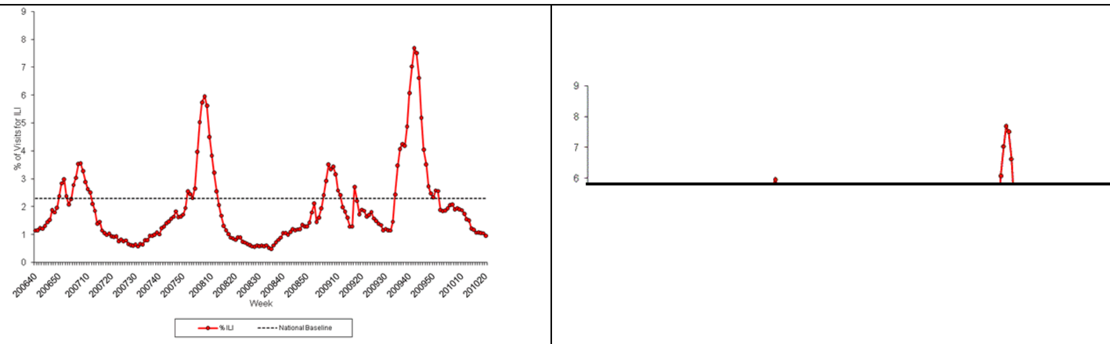


2(a). Earthquakes: the graph on the right is the lower 'dark-red' part of the graph on the left i.e. the earthquakes with magnitude more than 7 points in the Richter scale in large densely-populated areas in USA. The Y-axis on the left graph is the number of earthquakes over 7 Richter points, while on the right is number of earthquakes over 8 points.





2(b). Stock Markets: S&P 500, Returns (LogDiffs) of Weekly Closing Values [1/1/1990-1/2/2009]. The graph on the right is the upper part of the graph on the left – anything over the 0.02 points limit line, so the Peaks-Over-Threshold isolated points from the left full series. The Y-axis on both graphs is a derivation of squared differences of the index, and comes without units of measurement in practice and the literature.



2(c). Epidemics - Influenza : the graph on the right is the upper part of the graph on the left – anything over the seasonal influenza level of 6% points line (source: http://www.cdc.gov/h1n1flu/updates/us/). The variable measured is the percentage of Visits for Influenza-like Illness as reported from the United States Centers for Disease Control and Prevention. The Y-axis on the left graph is percentage of visits in monitored primary care units, while on the left is % of visits per day, threshold for visits over 6%.

Figure 2. Examples of an Intermittent Demand series as Peaks-Over-Threshold part of a regular time series

It is very important to forecast as accurately as passible for series such as these illustrated in figure 2. In the left part of figure 2(a), it is those big earthquakes that are potentially responsible for major catastrophes including tsunamis and nuclear disasters like the recent Fukushima one. In the left part of figure 2(b), it is those periods in the Stock markets that are extremely volatile and thus arguably considered as periods of high uncertainty leading to events like the Black Friday. In the left part of figure 2(c), it is those periods where virus activity converts suddenly into epidemics and require substantial resource from the national health systems.

3.2 Implications for practice from forecasting peaks-over-threshold series.

Given that none of the intermittent demand forecasting methods will give you the exact timing of a forthcoming extreme event, but rather will provide a cumulative forecast over a long period or just a ratio of a 'demand per period estimate', then the question becomes what can practitioners use this forecast for?

For each of the aforementioned application areas, we hereafter discuss the main practitioner implications from achieving an accurate cumulative forecast:

- Earthquakes (catastrophes): even if there is knowledge that a major earthquake is going to hit a region, it is hard to decide to evacuate cities, still you can influence and legislate the way buildings are built and increase the awareness and readiness of the. Furthermore you can ensure there is enough capital on hold so as to cope with the aftermath of the disaster
- Healthcare (epidemics): if we have advance knowledge of a forthcoming epidemic or pandemic over a period of time, then we could proactively source resources, get temporal staff in time, prepare large quantities of vaccines and set the whole national health systems in alert.
- Stock Markets (market crush): such a cumulative forecast would be an estimate of the aggregate risk over a certain (long) period of time, and thus necessary hedging actions could be taken in advance.

3.3 Non-OR quantitative methods for forecasting special events

Academics and practitioners do not use intermittent demand methods for forecasting special events of peaks-over-threshold time series. Instead, the current doctrine when forecasting in such a context is to use advanced probabilistic models. These methods typically require a lot of data and aim to reconstruct the distribution of the underlying phenomena. These methods come with a merit but definitely with a lot of caveats and constraints as well: big datasets needed, increased computational time, high mathematical complexity.

From a practical point of view, it is also very difficult to explain clearly to practitioners how these methods do actually work and the *acceptance of such models in industry and policy is far more challenging*. Practitioners want 'ownership' of the tools they use, and feel more comfortable with method that can understand how they work.

For the sake of completeness, we do provide hereafter make a short discussion of these methods. Extreme Value Theory (EVT) is dealing with the extreme deviations from the median of probability distributions, in order to estimate the probability of events that are more extreme than anything observed in the historic data in the past. Two approaches mainly exist:

- Deriving max/min series as an initial first step, that is generating an "Annual Maxima/Minima Series" (AMS) usually leading to the use Generalized Extreme Value Distribution as the most appropriate distribution to be fitted. Very often the number of relevant random events within a year usually is quite limited if any. Furthermore, very often analyses of observed AMS data lead to other distributions.
- Isolating the values of a normal series that exceed a threshold, usually referred to as "Point Over Threshold" (POT) method and yet again can lead to a few only or no values at all being extracted in any given year. The analysis involves fitting one distribution for the number of events in a basic time period (typically Poisson) and a second distribution for the size of the resulting POT values (typically a Generalized Pareto Distribution).

This latter approach clearly resembles in principle the logic of the Croston decomposition (1972), however Croston's approach is not trying to reconstruct the distributions of the two independent series (inter-demand arrivals and demand volumes), rather than just forecast the two independent series with simple and very robust exponential smoothing approaches, and then combine those two forecasts into a ratio forecast for the future demand per period.

4. Improving the performance of intermittent demand methods through a temporal aggregation approach

We have argued extensively so far for the potential of the use of intermittent demand forecasting methods for forecasting special events. This is not a paper about a specific method or a specific model. It is a paper making the argument that intermittent demand forecasting methods – as developed over the last 50 years in OR/MS journals – are very important. Far more important than just for spare parts forecasting and inventory management, that originally have been proposed for. These methods can be used for any Peaks-Over-Threshold series – and these can be derived (as illustrated in this article) from any series within any application area.

Moreover, what if the proposed intermittent demand methods in the literature could be become further improved? Then we would have and even more tempting proposition argued in this article. From this point on, we will discuss and existing methodology (Nikolopoulos et al., 2011), based on temporal aggregation, that is independent of forecasting models, and has been found to improve the performance of intermittent demand forecasting methods. The same is the theoretical and empirical evidence for fast demand forecasting methods (Rostami-Tabar et al., 2013; Spithourakis et al., 2011). In theory any method, statistical or machine learning, can been improved if applied in a different temporal frequency than the one for the original data, or through a combination of temporal frequencies (Athanassopoulos et al. 2017; Kourentzes et al., 2017).

We can further advocate that Naive is an appropriate approach when intermittence is that high as in the illustrated examples (Petropoulos et al., 2014; Syntetos et al., 2015). Naïve is the simplest forecasting method, very popular in industry and in fiancé (for short term financial returns), where the forecast for the next perio equals to the last known actual data point.Furthermore, it has been proven empirically that Naïve works better through the Aggregate-Disaggregate-Intermittent-Demand-Approach (ADIDA) non-overlapping temporal aggregation method-improving framework (Nikolopoulos et al., 2011).

In the ADIDA methodology, we start applying a method instead of the original frequency into many lower frequencies; see figure 3 for the numerical calculation and figure 4 for the respective graph. In figure 3 we start by monthly data but then we apply the method into quarterly and semester data. We could continue to years and beyond assuming sufficient number of observations. Any method could be applied thought the process but the sake of simplicity and without loss of generality here, we do apply Naïve. Nikolopoulos et al. (2011) used Naïve and SBA, Rostami-Tabar et al. (2013) use AR1 and MA1, Spithourakis et al. (2011) used Simple Exponential Smoothing (SES - Brown, 1963; Brown, 1956), Holt Exponential Smoothing (HES - Holt, 1957), Dampen Exponential Smoothing (DES - Gardner and McKenzie, 1985) and the Theta method (Assimakopoulos and Nikolopoulos, 2000).

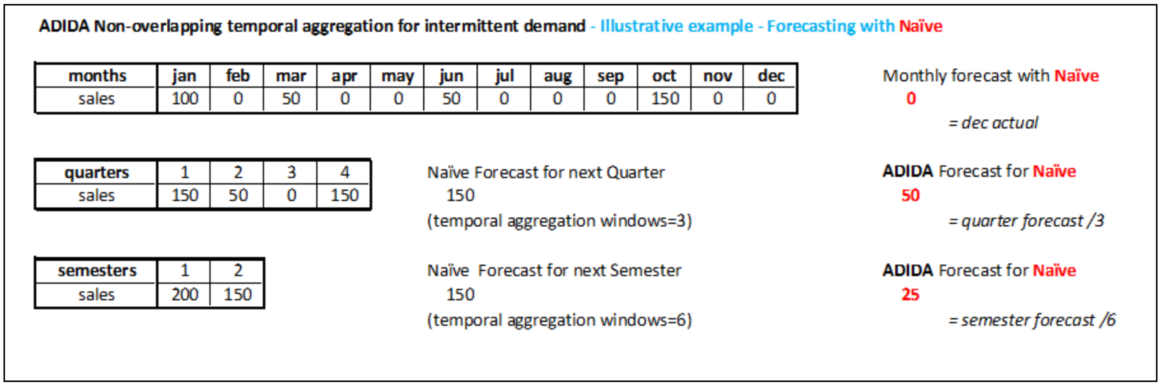


Figure 3. Monthly Sales in units: ADIDA Non-overlapping temporal aggregation for intermittent demand. Illustrative numerical example when Forecasting with the Naïve method.

In figure 3, we have a monthly series quite intermittent with four periods with demand over twelve periods overall. The intermittence is 0.5 as we have four periods with demand over six periods without demand.

- We first apply the Naïve methods on the original frequency, the monthly data. The last known demand, that of December, equal to 0, is the forecast for the immediate next January.
- We then construct a quarterly series through a non-overlapping temporal aggregation approach: that means that we add up the demand for January, February and march to get the first quarter. Then April, May and June for the second quarter. If we were doing overlapping aggregation the second quarter would be instead the sum of February, March and April (Boylan and Babai, 2016). We then apply the Naïve method on quarterly data and the forecast

for the next immediate quarter is that of the last known quarter equal to 150. Then the monthly forecast for each of the three months in that quarter is 150/3 =50 and thus this is the forecast for the immediate next January. This latter step assumes equal weights through the quarter and despite other propositions been made over the years, this is a very robust way to go about in the disaggregation phase (Nikolopoulos et al. , 2011)

- We then construct a semester series: that means that we add up the demand for first six months to get the first semester and the last six month sfor the second semester. We then apply the Naïve method on semester data and the forecast for the next immediate semester is that of the last known semester equal to 150. Then the monthly forecast for each of the six months in that semester is 150/6 =25 and thus this is the forecast for the immediate next January.

We always do evaluate the accuracy on the originally frequency and that is why we always prepare a monthly forecast in this example. If we wanted to evaluate the accuracy we should have kept some months hidden out as a holdout and that is what we do in the detailed empirical evaluation in the next section. A visual illustration of this numerical illustrative example (figure 3) is presented in figure 4.

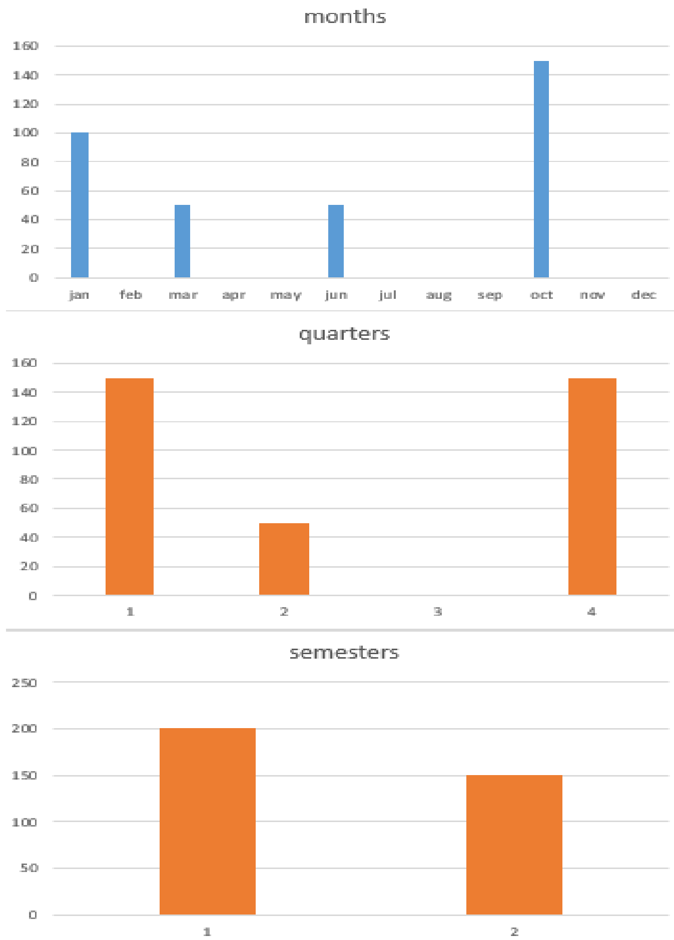


Figure 4. Monthly Sales in units: Illustrative visual example of non-overlapping temporal aggregation (ADIDA).

5. Simulations on real data

We have created an exhaustive time series of daily returns for the NASDAQ index, Daily Returns (=100%*(CP_t-CP_{t-1})/CP_{t-1}), since the origination of the index in 5th of Feb 1971 till the 13th of June 2014, totaling more than 10500 Daily Returns from the respective trading days for a period of 44 years (figure 5).

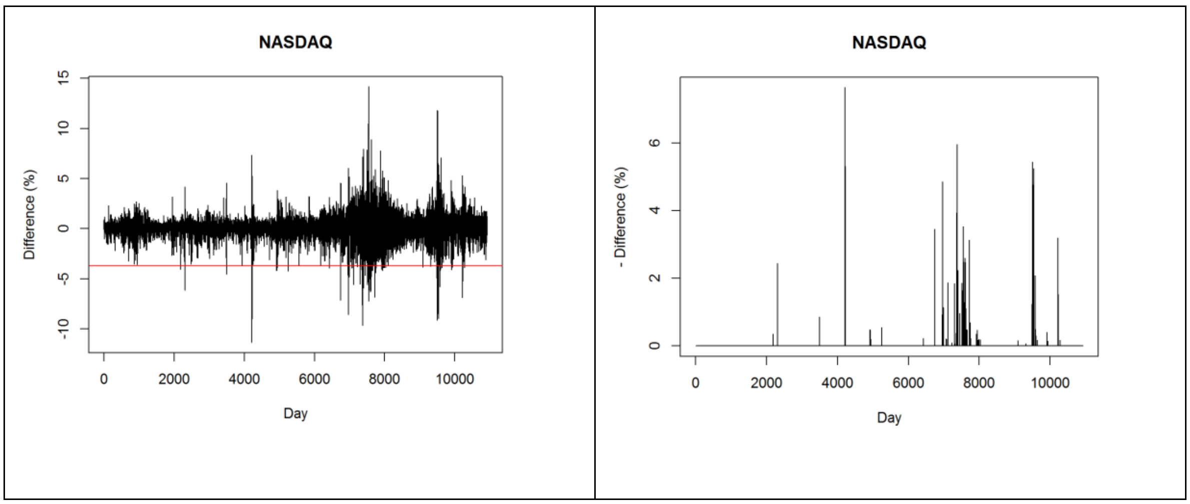


Figure 5. NASDAQ Daily returns from 5th Feb 1971 to 13th June of 2014: Negative Daily returns of more than 4% are illustrated on the right part of the graph. This is a returns series so the Y-axis is percentages (%)

We then create the Peaks-Over-threshold intermittent series for the data points under a negative threshold. This includes the daily returns that are of high risk as we lose a lot of money in these days. We select the days where the negative return is over 5 standard deviations from the mean in the left tail of the distribution, that sets a threshold of -4%. The resulting series (mirrored) is illustrated on the right part of figure 4.

This is an intermittent demand series, a series illustrating all the days in the market where significant amounts of money were lost and as such reflect very risky days. A cumulative forecast in such a context would proxy aggregate risk over a long period of time and this is exactly what we will try to forecast.

5.1 Forecasting

We create forecasts for 2 calendar weeks ahead (10 trading days) in the intermittent series of figure 4. We apply the Naive forecasting method over the ADIDA temporal aggregation framework and measure the average MAE over a 10-day window for various temporal aggregation windows ranging from 1 (original frequency) to 500 days.

So we first create series of lower frequencies aggregating every 2,3,4, …, 500 days and then we provide forecasts for the next immediate 10 days, and so on and so forth. Then we calculate the out-of-sample forecasting errors in these respective frequencies. This is illustrated in figures 6 and 7, where in the latter the first 100 aggregation windows are zoomed in so as to see clearly the reduction in the forecasting error; and thus the increase of the forecasting performance of the Naïve method.

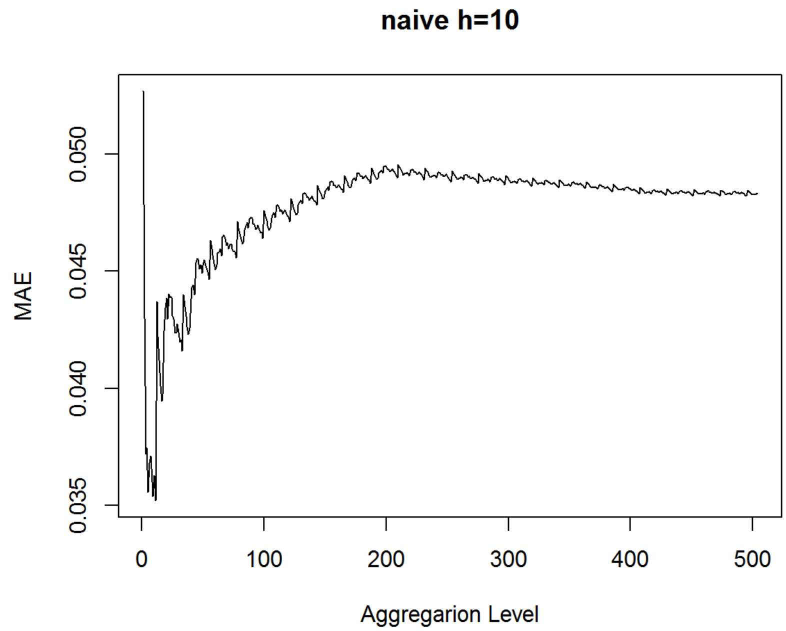


Figure 6. NASDAQ Negative Daily returns (over a threshold of -4%) forecasted with ADIDA(Naive). The graph illustrates the evolution of MAE over a 10-day period along various temporal aggregation windows (from 1 day to 500 days).

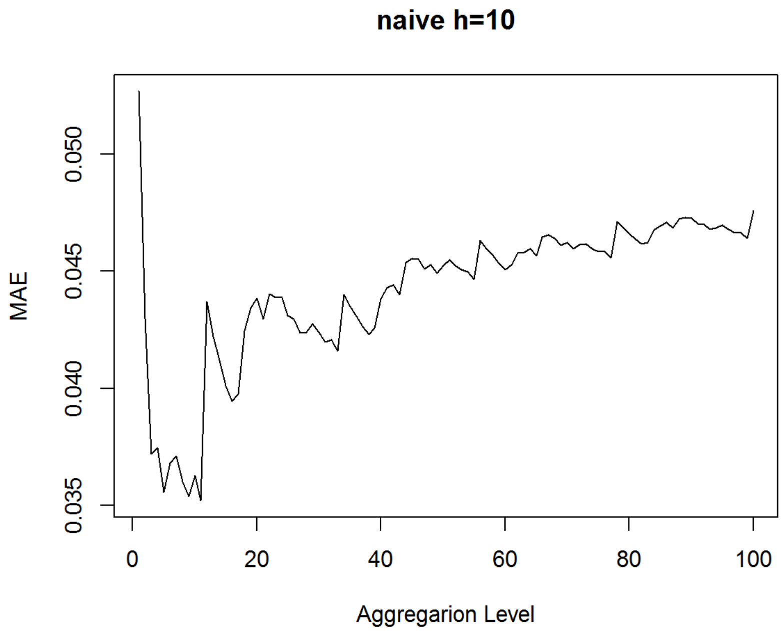


Figure 7. NASDAQ Negative Daily returns (over a threshold of -4%) forecasted with ADIDA(Naive). The graph illustrates the evolution of MAE over a 10-day period along various temporal aggregation windows

[Zoomed for a max of a window of 100 days]

As the temporal aggregation window becomes equal to 2, the error falls from over 0.05 to 0.037. Then as the temporal aggregation window increases to 3,4,5,6,7,8,9 we see local fluctuations of the error in the area of 0.035 to 0.037 up to the (empirically decided) minimum error that is achieved for an aggregation window of 10 days, that in this instance happens to match the target forecasting horizon. Thus, we can claim that the forecasting performance of the Naive method is improved as the temporal aggregation window increases until the global optimum is reached.

That means that the Naïve method maximizes its predictability potential when used in temporal aggregation windows of 10 days: so first we aggregate the data in an non-overlapping way every 10 days (two weeks' worth of trading days) and then forecast on that series; and that is how we do achieve maximum forecasting accuracy in this specific instance. This also implies that we better take decisions on the same window, so open investment positions in the market and close them after two weeks – rather than every day or any other period, as that is the window that maximizes our multivariate objective function: to forecast adequately (accuracy) and anticipate extremes (uncertainty). We will elaborate further on this latter insight in the following section.

5.1 Foresight

In a foresight context, we illustrate long-term-forecasting for 2 calendar years ahead (504 trading days). We apply two other methods – for the sake of generalization of our results – SBA (Syntetos and Boylan, 2001) and SES (Brown, 1963; Brown, 1956) yet again over the ADIDA temporal aggregation framework. We do respectively measure the average MAE over a 504-day window. This is illustrated in figures 8 (SBA) and 9 (SES).

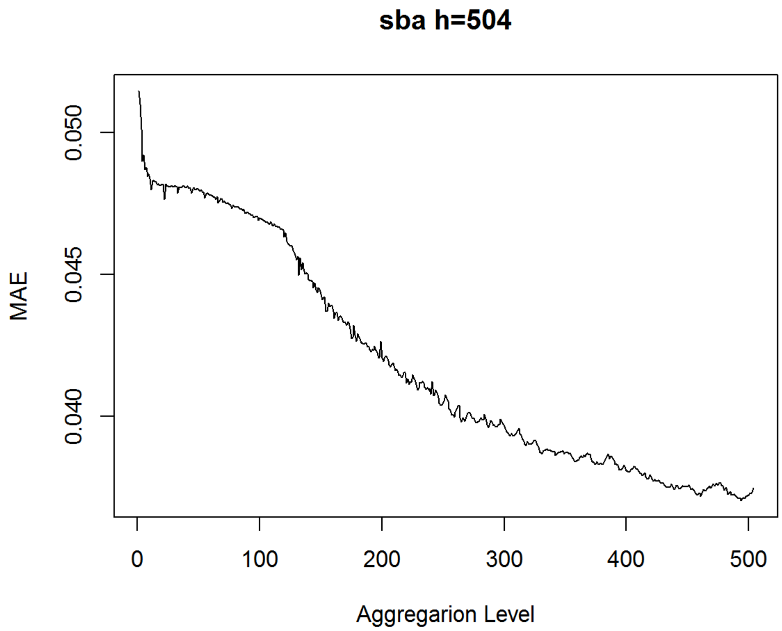


Figure 8. NASDAQ Negative Daily returns (over a threshold of -4%) forecasted with ADIDA(SBA). The graph illustrates the evolution MAE over a 504-day period along various temporal aggregation windows.

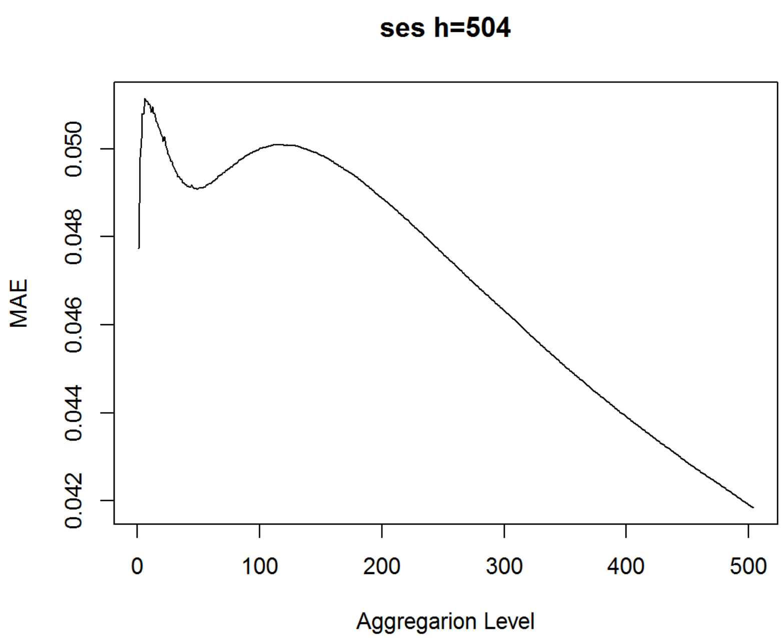


Figure 9. NASDAQ Negative Daily returns (over a threshold of -4%) forecasted with ADIDA(SES). The graph illustrates the evolution MAE over a 504-day period along various temporal aggregation windows.

In figure 8 and 9 is evident that the forecasting performance of both SBA and SES keeps on improving as the temporal aggregation window increases, and an empirical minimum error does not seem to be achieved. Nevertheless, the performance for SBA is monotonically becoming better as the temporal aggregation window increases, and is always better than the performance in the original frequency. While for SES working on any and aggregation window over 250 would always give better performance than working on the original frequency with SES. That would practically mean aggregating one-year’s worth of data, and then with SES providing a forecast for two years ahead.

Our simulation was not extended for aggregation of more than 504 trading days and as such there is still chances that there was an optimum level of aggregation to be identified.

6. Strategic planning in the presence of special events

In a forecasting task, the *forecasting horizon* is usually set and given a priori. This may be imposed by our customers, our suppliers, or any other stakeholder. This is meant to be serving a purpose and a decision linked to it. Sometimes it is even imposed by our IT systems. However, is this forecasting horizon, same or close by to the frequency where our predictability is maximized? Is this forecasting horizon,the one that minimizes the uncertainty around our forecasts and respective decisions?

If we follow further this argument, we may be able to rethink the forecasting/foresight horizon that we use in our applications. In other words we could:

"*Re-think the strategic planning horizon*".

What if we match it to the forecasting/foresight horizon that minimizes the forecasting error and thus the respective uncertainty around our forecasts and the future in general. This becomes more of an important and interesting question, if we are in a context of high uncertainty, like in the presence of special events like black and gray swans, where then we really 'battle against uncertainty' and any more improvement in forecasting accuracy could be proven critical.

One way to do this is by first creating the peaks-over-threshold series from our data of interest. Then we run an exhaustive simulation across: a) all intermittent demand forecasting methods, b) all forecasting horizons, c) all non-overlapping temporal aggregation levels (and even possibly all forecasting errors). Then we measure the forecasting error[9] we achieve. This is a 4D optimization problem and an attempt to visualize that is presented in figure 10.

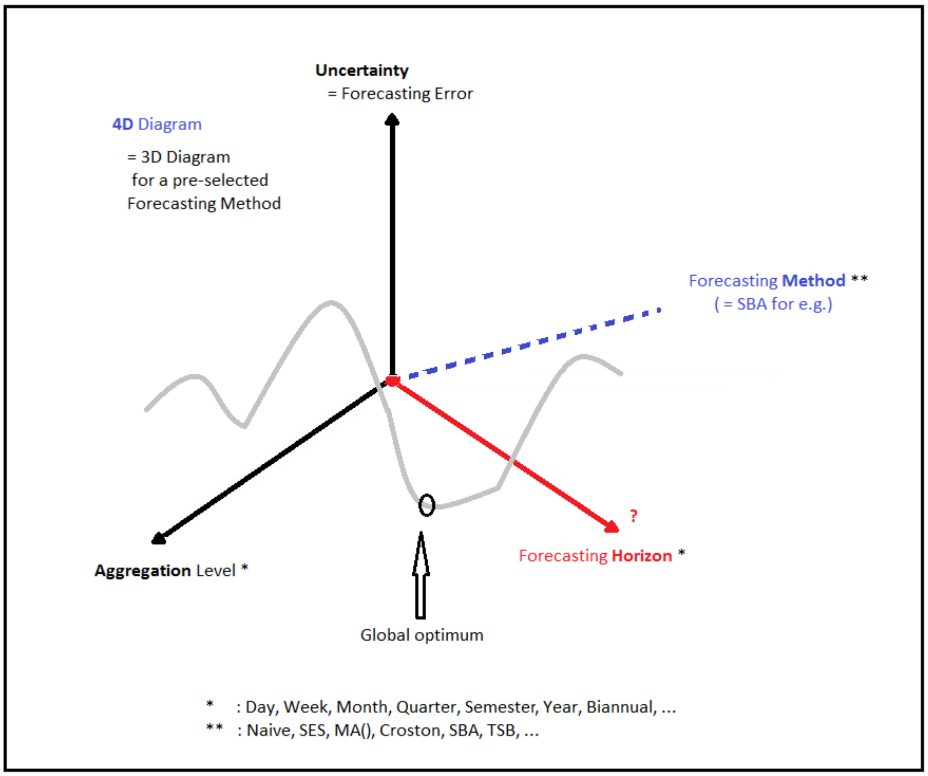


Figure 10. Re-thinking of the Strategic Planning horizon: what if we match it to the forecasting/foresight horizon that minimizes the forecasting error. This is a 3D snapshot (with Forecasting Method = SBA) of a 4D graph (for all possible methods as a 4th dimension - illustrated by the dotted-blue line).

[9] MAE or even better a 'super-metric' ranking across many accuracy metrics: MAE, MASE, MAPE, MSE, ME (Hyndman and Koehler, 2006).

If there is a forecasting horizon in which the forecasting accuracy is improved and thus uncertainty decreasing – a local or global optimum in the 4D space – then this is the level at which we maximaze our predictability. Whether or not that horizon is close-enough to the one we are using at the moment for strategic planning, is a different story. We may (and probably should) wish to consider switching our decision and strategic planning horizon to that one that is prescribed from figure 10. This will take a lot of organizational change and management persuasion, but the benefits will be clear from the improved forecasting accuracy and respective efficient decisions driven from it.

7. **Conclusions and future research**

In this research, we advocated for the broader use of OR forecasting tools; and in specific for intermittent demand forecasting methods for forecasting special events like black and gray swans. Our proposition is simple, fast, and a robust alternative to more complex non-OR models coming from econometrics. Furthermore, in a foresight context, we argue for a novel way of deciding the strategic planning horizon for phenomena prone to appearance special events.

Our contribution to the theory and practice of OR is dual:

a) First we empahtically argue for augmenting the use outside OR of a class of methods that meant to be primarily used for spare parts forecasting and inventory management.

b) Secondly we do not attempt to show superiority of our approach over existing approaches for forecasting and stragic planning: EVT, and 'managerially imposed forecasting horizons' respectively. We do however emphasize the *applicability*, *simplicity, speed* and *robustness* of our proposition, even in the (potential) expense of some optimality.

In any case, intermittent demand forecasting method should always be used as benchmarks in these aforementioned contexts.

We also believe that our proposition is intuitively appealing to practitioners as well. In situations of high uncertainty that do appear in all kinds of applications, our OR intermittent demand forecasting arsenal should be employed, even if it is for merely providing the benchmarks against more complex econometric approaches. Do not be surprised if simplicity prevails in this instance as well (Zellner et al., 2002).

For the future, we need to compare this simple family of methods with more complex methods that are coming from econometrics. The latter methods also do require far more data in order to be tuned to the task. Our proposition can work even with a minimal training set. We also need to investigate the appetite for change of the stakeholders, given organizations tend to be well tied to their decision plans and respective time horizons.

References

Assimakopoulos, V. and Nikolopoulos, K. (2000). "The Theta Model: A Decomposition Approach to Forecasting", *International Journal of Forecasting* 16 (4): 521-530.

Altay, N., Litteral, L.A., & Rudisill, F. (2012). Effects of correlation on intermittent demand forecasting and stock control, *International Journal of Production Economics* 135(1), 275–283.

Altay, N., Rudisill, F. & Litteral, L.A. (2008). Adapting Wright's modification of Holt's method to forecasting intermittent demand, *International Journal of Production Economics* 111(2), 389–408.

Athanasopoulos, G., Hyndman, R. J., Kourentzes, N. and Petropoulos, F. (2017). "Forecasting with temporal hierarchies", *European Journal of Operational Research* 262(1): 60-74.

Babai, M.Z., Ali, M. and Nikolopoulos, K. (2012) "Impact of Temporal Aggregation on Stock Control Performance of Intermittent Demand Estimators: Empirical Analysis", *OMEGA: The International Journal of Management Science* 40: 713–721.

Boylan, J. E. & Babai, M. Z. (2016). "On the performance of overlapping and non-overlapping temporal demand aggregation approaches", *International Journal of Production Economics* 181: 136-144.

Bozos, K. and Nikolopoulos, K. (2011) "Forecasting the Value Effect of Seasoned Equity Offering Announcements", *European Journal of Operational Research* 214: 418–427.

Brown, Robert G. (1956). *Exponential Smoothing for Predicting Demand. Cambridge*, Massachusetts: Arthur D. Little Inc.

Brown, Robert Goodell (1963). *Smoothing Forecasting and Prediction of Discrete Time Series*. Englewood Cliffs, NJ: Prentice-Hall.

Croston, J.D. (1972). "Forecasting and stock control for intermittent demands". *Operational Research Quarterly* 23: 289-303.

Davenport, T. H. & Harris, J. G. (2007). *Competing on Analytics: The New Science of Winning*, Harvard Business School Press Boston, MA

Fildes, R., Nikolopoulos, K., Crone, S., and Syntetos, A. (2008) "Forecasting and Operational Research – a Review". *Journal of the Operational Research Society* 59: 1150–1172.

Fildes, R. and Nikolopoulos, K. (2006) "Spyros Makridakis: An Interview with the International Journal of Forecasting". *International Journal of Forecasting* 22(3): 625-636.

Gardner Jr., E.S. & McKenzie, E. (1985). "Forecasting trends in time series", *Management Science* 31, 1237-1246.

Holt, Charles C. (1957). "Forecasting Trends and Seasonal by Exponentially Weighted Averages". *Office of Naval Research Memorandum*. 52. reprinted in Holt, Charles C. (January–March 2004). "Forecasting Trends and Seasonal by Exponentially Weighted Averages". *International Journal of Forecasting*. 20 (1): 5–10

Hyndman, R. J. & Koehler, A. B. (2006). “Another look at measures of forecast accuracy”,*International Journal of Forecasting* 22(4): 679-688.

Katsagounos, I., Nikolopoulos, K. & Thomakos, D. D. (2020) “Looking for the needle in the haystack: Evidence of the superforecasting hypothesis when time and samples are limited”, *European Journal of Operational Research*, First round: Major Revision

Kourentzes, N., Rostami-Tabar, B. and Barrow, D. K. (2017). “Demand forecasting by temporal aggregation: Using optimal or multiple aggregation levels?”, *Journal of Business Research* 78: pp. 1-9.

Kourentzes,N., Petropoulos,F., & Trapero, J.R. (2014). “Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting* 30(2), 291–302.

Kremer, M., Siemsen, E. and Thomas, D. J. (2016). “The Sum and Its Parts: Judgmental Hierarchical Forecasting”, *Management Science* 62(9): 2457-2764.

Leadbetter, M.R. (1991). "On a basis for 'Peaks over Threshold' modeling", *Statistics & Probability Letters* 12 (4): 357–362.

Makridakis, S., Spiliotis, E. & Assimakopoulos, A. (2018) “The M4 Competition: Results, findings, conclusion and way forward”, *International Journal of Forecasting* 34(4), 802-808.

Nikolopoulos, K. and Thomakos, D. D. (2019), *Forecasting with the Theta Method: Theory & Applications*, Wiley: New Jersey.

Nikolopoulos, K. and Petropoulos, F. (2018) "Forecasting for big data: does suboptimality matter?", *Computers and OR* 98: 322-329.

Nikolopoulos, K., Babai, M. Z., and Bozos, K. (2016) “Forecasting Sporadic demand with Nearest Neighbor approaches”, *International Journal of Production Economics* 177: 139–148.

Nikolopoulos, K., Buxton, S., Khammash, M., and Stern, P. (2016) “Forecasting Branded and Generic Pharmaceuticals”, *International Journal of Forecasting* 32(2): 344–357.

Nikolopoulos, K., Syntetos, A.A., Boylan, J., Petropoulos, F., and Assimakopoulos, V. (2011). "An aggregate-disaggregate intermittent demand approach (ADIDA) to forecasting: an empirical proposition and analysis". *Journal of the Operational Research Society* 62: 544-554.

Nikolopoulos, K., Goodwin, P., Patelis, A. and Assimakopoulos, V. (2007), "Forecasting with cue information: a comparison of multiple regression with alternative forecasting approaches". *European Journal of Operational Research* 180(1): 354-368.

Nikolopoulos, K., Petropoulos, F., Sanchez-Rodrigues, V. Pettit, S. and Beresford, A., (2019) “A risk-mitigation model driven from the level of forecastability of Black Swans: prepare and respond to major Earthquakes through a dynamic Temporal and Spatial Aggregation forecasting framework” in *Bangor Business School Working Paper series*, BBSWP/19/17 : 1-17.

Petropoulos, C., Nikolopoulos, K. Patelis, A., Assimakopoulos, V. and Askounis, D. (2006). “Tourism Technical Analysis System”. *Tourism Economics* 12 (4): 543-563.

Petropoulos, F., Kourentzes, N., Nikolopoulos, K. and Siemsen, E. (2018) “Judgmental Selection of Forecasting Models”, *Journal of Operations Management* 60: 34-46.

Petropoulos, F., Kourentzes, N. and Nikolopoulos, K. (2016) "Another look at estimators for intermittent demand", *International Journal of Production Economics* 181: 154–161.

Petropoulos, F., and Kourentzes, N. (2014). "Forecast combinations for intermittent demand". *Journal of the Operational Research Society*. 66, 914-924

Petropoulos, F., Makridakis, S., Assimakopoulos, V. and Nikolopoulos, K. (2014). "'Horses for Courses' in demand forecasting", *European Journal of Operational Research* 237 (1): 152-163.

Rostami-Tabar, B., Babai, M. Z., Ali, M. and Boylan, J. (2019). “The impact of temporal aggregation on supply chains with ARMA(1,1) demand processes”, *European Journal of Operational Research* 273(3): 920-932.

Rostami-Tabar, B., Babai, M. Z., Syntetos, A. and Ducqet, Y. (2013). “Demand forecasting by temporal aggregation”. *Naval Research Logistics* 60(6): 479-498.

Spithourakis, G.P., Petropoulos, F., Babai, M.Z., Nikolopoulos, K. and Assimakopoulos, V., (2011). “Improving the performance of popular supply chain forecasting techniques: an empirical investigation”, *Supply Chain Forum: an international journal* 12(4): 16-25.

Syntetos, A.A. and Boylan, J.E. (2001). "On the bias of intermittent demand estimates2. *International Journal of Production Economics* 71 (1-3): 457-466.

Syntetos, A.A., Boylan, J.E, and Croston, J.D. (2005). "On the categorization of demand patterns". *Journal of the Operational Research Society* 56: 495-503.

Syntetos A.A., Babai M.Z. and Gardner, E.S. Jr. (2015). "Forecasting Intermittent Inventory Demands: Parametric Methods vs. Bootstrapping". *Journal of Business Research*, forthcoming

Syntetos, A., Babai, Z.M., Boylan, J., Kolassa, S. and Nikolopoulos, K. (2016). “Supply Chain Forecasting: Theory, Practice, their Gap and the Future”, *European Journal of Operational Research* 252 (1): 1-26.

Perera N, Hurley J, Fahimnia B, and Reisi Ardali M (2019). “The Human Factor in Supply Chain Forecasting: A Systematic Review”, *European Journal of Operational Research* 274 (2): 574-600

Punia, S., Nikolopoulos, K., Singh, S. P., Madaan, J. K. (2019). “Deep learning with long short-term memory networks and random forests for multi-channel retail demand forecasting”, *Internationla Journal of Production Research*, under review

Solari, S., Eguen, M., Polo, M. J. and Losada, M. A. (2017), Peaks Over Threshold (POT): A methodology for automatic threshold estimation using goodness of fit p-value, *Water Resources Research* 53: 2833–2849.

Taleb, N.N. (2007), *The Black Swan: The Impact of the Highly Improbable*. New York: Random House and Penguin.

Tetlock, P. & Gardner, D. (2015). Superforecasting: The Art and Science of Prediction Paperback, Crown;

Teunter, R.H., Syntetos, A.A., & Babai, M.Z. (2011). Intermittent demand: linking forecasting to inventory obsolescence. *European Journal of Operational Research* 214, 606–615.

Willemain, T.R., Smart, C.N., and Schwarz, H.F. (2004). "A new approach to forecasting intermittent demand for service parts inventories". *International Journal of Forecasting* 20: 375-387.

Zellner, A., Keuzenkamp, H. A. & McAleer, M. (2002). *Simplicity, Inference and Modelling: Keeping It Sophisticatedly Simple*, Cambridge University Press.

Authors Bio

Professor Konstantinos (Kostas) Nikolopoulos is the Director of forLAB www.forlab.eu . He is an expert in Time Series Analysis & Forecasting, Decision Support Systems and Management Judgment. Professor Nikolopoulos received his Engineering Doctorate in 2002 from National Technical University of Athens. He has worked in the past for the University of Manchester and other prestigious institutions in U.K. and Greece, and now holds the Chair in Business Analytics in Bangor Business School. His work has appeared in JOM, EJOR, IJF, JoF, JORS, Omega, IJPE, IJPR, IMA JMM, JBR, JCIS, IMDS, AE among other journals, and he is an Associate Editor of Oxford IMA Journal of Management Mathematics and Supply Chain Forum: An International Journal. He is co-originator of the Theta forecasting method and the ADIDA temporal aggregation method-improving framework.