# 'Horses for Courses' in Demand Forecasting

Fotios Petropoulos[1], Spyros Makridakis[2], Vassilios Assimakopoulos[3], and Konstantinos Nikolopoulos[4,*]

[1]Lancaster Centre for Forecasting, Lancaster University, Lancaster, UK;
f.petropoulos@lancaster.ac.uk
[2]INSEAD Business School, France;
spyros.makridakis@insead.edu
[3]Forecasting & Strategy Unit, National Technical University of Athens, Greece;
vassim@fsu.gr
[4]Bangor Business School, Prifysgol Bangor University, Bangor, Wales, UK;
k.nikolopoulos@bangor.ac.uk

## Abstract

Forecasting as a scientific discipline has progressed a lot in the last forty years, with Nobel prizes being awarded for seminar work in the field, most notably to Engle, Granger and Kahneman. Despite these advances, even today we are unable to answer a very simple question, the one that is always the first tabled during discussions with practitioners: "what is the best method for *my data*?". In essence, as there are *horses for courses*, there must also be forecasting methods that are more tailored to some types of data, and, therefore, enable practitioners to make informed method selection when facing new data. The current study attempts to shed light on this direction via identifying the main determinants of forecasting accuracy, through simulations and empirical investigations involving fourteen popular forecasting methods (and combinations of them), seven time series features (*seasonality*, *trend*, *cycle*, *randomness*, *number of observations*, *inter-demand interval* and *coefficient of variation*) and one strategic decision (the *forecasting horizon*). Our main findings dictate that forecasting accuracy is influenced as follows: a) for fast-moving data, *cycle* and *randomness* have the biggest (negative) effect and the longer the *forecasting horizon*, the more accuracy decreases; b) for intermittent data, *inter-demand interval* has bigger (negative) impact than the *coefficient of variation*; c) for all types of data, increasing the length of a series has a small positive effect.

**Keywords**: Forecasting methods, Time series methods, Forecasting accuracy, M-Competitions, Simulation

* Corresponding Author

## 1. Introduction

Forecasts are important for all decision-making tasks, from inventory management and scheduling to planning and strategic management. Makridakis and Hibon (2000) advocated: "predictions remain the foundation of all science". To that end, identification of the best forecasting techniques for each data set, or, even, for each series separately, is still the 'holy grail' in the forecasting field, and, as a result, empirical comparisons to this direction are considered very important (Fildes and Makridakis, 1995). Advanced, sophisticated and simpler extrapolation methods could be associated with specific features of data. The development of a protocol for automatic selection of the best tools for resolving each problem, a protocol that would guarantee minimum out-of-sample forecasting error and therefore have a substantial impact on decision making, is the ultimate challenge for researchers and practitioners in the field.

As early as the late 1960s and most of the 1970s, several researchers (Kirby, 1966; Levine, 1967; Cooper, 1972; Naylor and Seaks, 1972; Krampf, 1972; Groff, 1973; Newbold and Granger, 1974; Makridakis and Hibon, 1979) sought to determine the accuracy of various forecasting methods in order to select the most appropriate one(s). In addition, psychologists have been concerned with judgmental predictions and their accuracy, as well as the biases that affect such predictions, for more than half a century (Meehl, 1954; Slovic, 1972; Kahneman and Tversky, 1973; Dawes, 1979; Meehl, 1986; Tversky and Kahneman, 1982; Hogarth, 1987). Amongst these biases, those affecting forecasting include over-optimism and wishful thinking, recency, availability, anchoring, illusory correlations and the underestimation of uncertainty. In a recent book, Kahneman (2011) describes these and other biases whilst also discussing what can be done to avoid, or minimize their negative consequences and emphatically states: "the research suggests a surprising conclusion: to maximize predictive accuracy, final decisions should be left to formulas, especially in low-validity environments" (Kahneman, 2011, p.225). Moreover, the growing demand for forecasting big data (e.g. more than 200,000 time series for major retailers) renders the use of automatic statistical procedures necessary.

The purpose of this study is to measure the extent to which each of seven time series features (*seasonality*, *trend*, *cycle*, *randomness*, *number of observations*, *inter-demand interval* and *coefficient of variation*) and one strategic decision (the

*forecasting horizon*) affect forecasting accuracy. In order to do this, we measure the impact of each of these eight factors[1] by generating a large number of time series - as well as using real data, and measuring the accuracy of the forecasts derived from fourteen methods and five combinations of them. Furthermore, a multiple regression analysis is performed to measure the extent to which each of the factors affects the accuracy of each of the time series methods/combinations. The findings of this research could be very useful for practitioners if used for the appropriate selection of the best statistical forecasting practices based on an ex-ante analysis of their data (and their respective features).

This paper is structured as follows: after the literature review (Section 2), the simulation design for fast-moving and intermittent demand data is discussed in Section 3. In Section 4 the accuracy results are presented. Section 5 discusses the findings and Section 6 presents the practical implications for decision makers. Finally, Section 7 concludes and suggests possible avenues for future research.

## 2. Background Literature

Extrapolation models are used very often when facing large amounts of data. Among them, exponential smoothing forecasting approaches were developed in the early 1950s and have become very popular amongst practitioners. Their main advantages are simplicity of implementation, relatively low computational intensiveness and no requirement for lengthy series, whilst being appropriate for short-term forecast horizons over a large number of items. Single Exponential Smoothing (SES - Brown, 1956) uses only one smoothing parameter and is forecasting quite accurately stationary data. Holt's two parameters approach (1957) expands the Single model with a smoothing parameter for the slope, making the method more appropriate for trended data. The Holt-Winters approach (Winters, 1960) is an expansion upon the Holt trended model, which assumes an additive or multiplicative seasonality in the data. Gardner and McKenzie (1985) added a dampening factor ($0<\varphi<1$) applied directly on the trend component, resulting in a very successful approach that is often considered the benchmark in many empirical evaluations. Assimakopoulos and Nikolopoulos (2000) proposed the Theta model - a prima facie variation of SES with drift, with the full theoretical underpinnings presented by Thomakos and

[1] We use the term 'factor' to refer to both the data features as well as the strategic decision (*forecasting horizon*) that we will examine their impact on forecasting accuracy through this study.

Nikolopoulos (2014), a method that topped the M3-Competition, the largest empirical forecasting competition to date (Makridakis and Hibon, 2000, Appendix B).

On the other hand, the more complex but quite popular Box-Jenkins methodology (Box and Jenkins, 1970) uses an iterative three-step approach (model identification, parameter estimation and model checking) in order to find the best-fit ARIMA model. To date ARIMA models are still considered the dominant benchmark in empirical forecasting evaluations, and find great popularity among OR researchers in applications spanning from hospitality and production to healthcare and climate forecasting (for e.g. see Broyles et al., 2010; Cao et al., 2012; Cang and Yu, 2014).

One result that stands for fast-moving data is that combining improves predictive accuracy (Surowiecki, 2005; Clemen, 1989; Makridakis and Winkler, 1983). In addition to this, combining reduces the variance of forecasting errors and therefore the uncertainty in predictions, rendering the selection of combinations less risky than individual methods (Hibon and Evgeniou, 2005). Many recent studies have verified that the combination of methods leads to more accurate forecasts, whilst, at the same time proposing more sophisticated weightings such as the trimmed and Winsorized means (Jose and Winkler, 2008), and the use of information criteria (Taylor, 2008; Kolassa, 2011).

For count data/intermittent data, Croston (1972) proposed decomposing the data into two subseries (demands and intervals) with Syntetos and Boylan (2005) proposing a bias-correction to the Croston method (Syntetos and Boylan Approximation or SBA). More recently, Teunter et al. (2011) suggested a decomposition method that relies on the separate extrapolation of the non-zero demands and the probability to have a demand. This method is very useful in cases of obsolescence. Lastly, simpler approaches, such as Naïve, Moving Averages and SES, have also been quite popular for such data especially among practitioners.

An interesting spin-off from the later intermittent demand literature came from Nikolopoulos et al. (2011) with the ADIDA non-overlapping temporal aggregation forecasting framework, that although designed and successfully evaluated empirically on count data (Babai et al., 2012), the implications pretty fast span out for fast-moving data as well (Spithourakis et al., 2011; Kourentzes et al., 2014). The proposed framework soon was perceived as a forecasting method "self-improving" mechanism that by changing the data series features through frequency transformation, can help extrapolation methods achieve better accuracy performance.

The first theoretical results for the ADIDA framework appeared recently in the literature (Spithourakis et al., 2013; Rostami-Tabar et al., 2013).

### 2.1 'Horses for courses'

Given the plethora of the aforementioned methods, it is now even more unclear: when should each method be used? Many researchers compared the performance of aggregate and individual selection strategies (Fildes 1989; Shah 1997; Fildes and Petropoulos, 2013). While selecting a single method for an entire data set would make sense for homogeneous data, model selection should be done individually (per series) when we deal with heterogeneous data, as to capture the different features met in each series.

Pegels (1969) presented the first graphical classification for exponential smoothing models, separating trend from cycle patterns, and also as additive from multiplicative forms. In a simulation study, Adam (1973) evaluated several forecasting models across five different demand patterns, including constant, linear trend, seasonal and step function. His findings indicate that no single model is consistently better than the others, and their performance depends primarily on the demand pattern, the forecasting horizon and the randomness, and secondarily on the selected accuracy metric. Gardner and McKenzie (1988) provided a procedure for model identification in the case of large forecasting applications. Their selected course of action involved the calculation of variances at various levels of differences in data, and using those for classifying the underlying pattern of the time series (constant or trended, seasonal or not seasonal, and so on).

A first attempt for a rule-based selection procedure of the best model derived from Collopy and Armstrong (1992). They proposed a framework that combines forecasting expertise with domain knowledge in order to produce forecasts based on the characteristics of the data. Their procedure consisted of 99 rules and four extrapolation techniques, while 18 time series features were used. A simplified domain knowledge-free version of this rule-based procedure was presented by Adya et al. (2000), using just 64 rules, three forecasting methods and six time series features. In order to render the procedure fully automated, Adya et al. (2001) presented an automatic identification of time series features for rule-based forecasting, which reduces significantly the forecasting cost of large data sets without serious losses in accuracy.

Shah (1997) proposed a seven-step model selection procedure for univariate series forecasting, using an individual selection rule based on 26 features. In a later study, Meade (2000) used 25 simple statistics as explanatory variables in order to predict the forecasting accuracy performance of nine extrapolation methods and, thus, select the most promising one. His results were evaluated on two empirical data sets.

A different path for model selection focuses on the application of families of methods (for example, Exponential Smoothing or ARIMA) and subsequently the selection of the single method that has the best trade-off between the goodness-of-fit and the complexity of the problem. To this end, the use of information criteria (Hyndman et al. 2002) has been very popular. At the same time, there is little to distinguish from the application of different information criteria (Billah et al., 2006). A disadvantage of model selection with information criteria is the inability to compare across different families of methods. An alternative to the use of information criteria is the evaluation of the performance of methods in a hold-out sample where forecasts are calculated for multiple origins and for single or multiple lead times (Fildes and Petropoulos, 2013). Depending on the specific experimental design, this strategy is known as "validation" or "cross-validation".

For count data, Bacchetti and Saccani (2012) provided a comprehensive literature review of the classification methods, whilst a demand-based classification for intermittent demand was proposed by Syntetos et al. (2005), later revised by Kostenko and Hyndman (2006).

Lastly, it is worth emphasizing that various similar attempts to identify suitable methods for forecasting cross-sectional data had been presented over the years; as for example in Nikolopoulos et al. (2007) in a marketing application and Bozos and Nikolopoulos (2011) in a strategic financial decision-making application, where a series of economics, econometrics, time series, artificial intelligence and computational intensive approaches as well as human judgment were compared within the context of the respective investigations.

### 2.2 Forecasting Competitions

Forecasting competitions have evaluated the performance of time series methods (Makridakis et al., 1982; Makridakis et al., 1993; Makridakis and Hibon, 2000) in order to better understand their relative accuracy and improve their usefulness. Since then, a large number of studies have compared the accuracy of various methods in

different forecasting settings. For example, Franses and van Dijk (2005) concluded that simpler models for seasonality perform better for short horizons. At the same time, more complex models should be preferred for longer forecasting horizons.

Furthermore, many researchers focused on the performance of exponential smoothing methods. Gardner (2006) compared the performance of damped trend to the class of state-space models. Using the data sets from the M and M3 forecasting competitions, he concluded that the damped approach is more robust and accurate than the individual selection of models through information criteria in almost every case, except for the short horizons of the monthly M3 data. Gardner and Diaz-Saiz (2008) explored the performance of exponential smoothing methods using telecommunications data. An analysis of the results suggested that SES with drift - a simplification of the Theta method (Assimakopoulos and Nikolopoulos, 2000) as proposed by Hyndman and Billah (2003) - provided the most accurate forecasts compared to any other smoothing method for every horizon.

Crone et al. (2011) conducted a forecasting competition for Computationally Intensive approaches, most notably Artificial Neural Networks (ANN). One of their main findings was that only one ANN method outperformed the damped trend. Lastly, Athanasopoulos et al. (2011) found that tourism data are best extrapolated using time series approaches rather than causal models, whilst the forecasting performance of "Naive" for annual data was "hard to beat". This result is also evidenced in another relevant study using tourism data (Gil-Alana et al., 2008), in which a simple model outperformed more complex ones for short horizons.

### 2.3 Research Questions

Having revisited all this literature, we believe there is still scope for studies investigating what makes some methods more (or less) accurate, and under what conditions; having that said, the main *Research Question (RQ)* of this study is as follows:

> RQ: *How various factors affect - if at all, the forecasting accuracy of time series extrapolation methods*?

To address this question, we design two extensive simulations for fast-moving and intermittent data respectively - as well as empirical evaluations in real data, involving fourteen univariate forecasting methods and five combinations of them.

Consequently, the extent of the influence of each factor is calculated through regression analysis.

## 3. Simulation of Data

### 3.1 Simulation of fast-moving data

Before we start elaborating on the empirical investigations, we need to formally introduce the data features that we will simulate in this study.

To that end we were inspired originally by the work of Adam (1973) that he identified the importance of *trend, seasonality, randomness* and the *forecasting horizon*. We were further influenced by the work of Collopy and Armstrong (1992) where they proposed their Rule-Based Forecasting (RBF) framework, the very essence of which is dominated by the identification of data featuresAmong the selected time series features, the *trend*, the *cycle*, *seasonality* and the *length* of the series were of key importance. Finally, the work of Nikolopoulos and Assimakopoulos (2003) where many of these data features were used as key elements in the object-oriented architecture of a prototype Forecasting Support System - TIFIS, gave us more firm evidence on the importance of the aforementioned data features.

The level of temporal aggregation (frequency) and the level of cross-sectional aggregation (level in hierarchy) of the data were not considered in this study. We focus on the forecasting performance of a specific level of temporal/cross-sectional aggregation. Most of the patterns described above may be observed at any level of aggregation and thus this latter feature was not simulated. The only exception is the seasonality. The effects of temporal and cross-sectional aggregation have been addressed elsewhere. Nikolopoulos et al. (2011) and the ADIDA framework indicate that forecasters may via temporal aggregations switch to different frequencies (than the ones the data are observed), while Kourentzes et al. (2014) present a framework to efficiently combine forecasts derived from multiple frequencies. Lastly, Athanasopoulos et al. (2009) explore different approaches to hierarchical forecasting, proposing an "optimal" approach, which provides reconciled forecasts at every level.

So the RQ can now be narrowed for fast-moving data as follows:

RQ1: *How six factors (seasonality, trend, cycle, randomness, the number of observations and the forecasting horizon) affect, if at all, the forecasting accuracy of time series extrapolation methods on fast-moving data*?

To generate simulated series for testing the above-mentioned research question, each of the first five factors (seasonality, trend, cycle, randomness and the number of observations) was varied around six levels (see Table 1) while 10,000 series were randomly generated at each level (ceteris paribus). Since there are six levels and five factors to vary, there is a total of 7,776 ($6^5$) combinations, resulting in 77,760,000 generated time series covering every possible combination. For each of the generated time series, 18 forecasts were produced. The values and variation of each of the six levels was selected by using the respective ranges of the 1,428 real monthly series of the M3 competition.

**Table 1:** Levels for the five factors for fast-moving data

| **Components** | **Level** | | | | | |
|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** |
| Seasonality | 0.0 | 0.5 | 2.5 | 7.5 | 13.5 | 20.0 |
| Trend | 0.0 | 0.6 | 1.2 | 1.8 | 2.4 | 3.0 |
| Cycle | 0.0 | 0.4 | 0.8 | 1.2 | 1.6 | 2.0 |
| Randomness | 0.0 | 0.5 | 2.0 | 4.0 | 7.0 | 10.0 |
| Number of observations | 36 | 48 | 60 | 84 | 108 | 144 |

The generation procedure assumes a deterministic, multiplicative model where each component is applied individually as suggested by Miller and Williams (2003), but in addition we introduce a cycle component as well. Thus:

$$X_t = S_t \cdot T_t \cdot C_t \cdot R_t \quad (1)$$

where $X_t$ is the series, $S_t$ is the seasonal component, $T_t$ is the trend component, $C_t$ is the cycle component and $R_t$ is the random component.

The procedure of simulating fast-moving series is described below. First, a vector of length equal to a selected *number of observations* plus 18 (out-of-sample) is defined, with all values being set equal to a randomly selected initial level (*L*). This vector is multiplied by the respective seasonal indices, which, given a *seasonality* level (*SL*), are defined as:

$$S_t = SI_k * \left( \frac{SL}{MAP(FD)} \right) \quad (2)$$

where *SI* is a zero-based single dimensional array of 12 values, containing the mean seasonality curve of the monthly seasonal time series of M3-Competition and *MAP*(*FD*) represents a normalization factor calculated as the mean absolute percentage of first differences of the *SI* values. Then, the *trend* component is applied to each observation. This component is equal to:

$$T_t = \sqrt[p]{(1+TL)^t} \tag{3}$$

where *TL* is the selected trend level and *p* represents the number of periods within a year. The cyclical component, $C_t$, is introduced as:

$$C_t = C_{t-1} + N_t(CL, CV = 1/3) \tag{4}$$

where $C_0 = 0$ and $N_t$ is a normally distributed random variable with mean value *CL* (the selected level of the *cycle* component) and a standard deviation so that CV=1/3. A new value of $N_t$ is generated for each data point. Lastly, for the randomness component, a normally distributed and randomly selected variable with mean value *RL* (level of *randomness*) and CV=1/3 is generated, or:

$$R_t = N_t(RL, CV = 1/3) \tag{5}$$

Lastly, $\boldsymbol{X_t}$ is calculated as follows:

$$X_t = L \cdot S_t \cdot T_t \cdot C_t \cdot R_t \tag{6}$$

The forecasting methods used in the study are (for more details on these methods, see appendix A):

**Naïve 1**, **Naïve 2**, four exponential smoothing methods (**Single, Holt, Damped, Holt-Winters**), **Theta** (Assimakopoulos and Nikolopoulos, 2000), **Linear Trend** and two commercial packages (**Autobox** and **Forecast Pro**).

Moreover, the following five combinations of the above methods were constructed:

- Single-Damped (**SD**)
- Single-Holt-Damped (**SHD**)
- Single-Theta (**ST**)
- Single-Damped-Theta (**SDT**)
- Single-Holt-Damped-Theta (**SHDT**)

With the six methods (Naïve 2, Single, Holt, Damped, Linear Trend and Theta) not suitable to handle seasonality, the forecasts were produced following a three-step procedure: firstly the data was deseasonalized via a Classical Decomposition approach. Secondly, 18 forecasts were computed using the

deseasonalized data. Finally, these 18 forecasts were reseasonalized using the same seasonal indices as the Classical Decomposition. The remaining methods (Naïve 1 and Holt-Winters) and the methods implemented in the commercial packages (Autobox and Forecast Pro) where applied directly to the original data. The selection of the optimal forecasting model was always carried out without using the last 18 out-of-sample observations which were being kept for evaluating the forecasting accuracy of each method. For all forecasting methods, except the two commercial packages, 10,000 series were generated for each of the 7,776 permutations; in contrast, for the two commercial packages only 300 series were generated for each of the 7,776 permutations (i.e. 2,332,800 series in total) because of the time required in order to run them.

The forecasting accuracy was measured by comparing the 18 hold out data points with the 18 point forecasts. Three accuracy metrics were calculated:

- The Symmetric Mean Absolute Percentage Error (sMAPE), the main metric of the M3 competition (Makridakis and Hibon, 2000).
- The Percentage Better, where the accuracy of each method was benchmarked against Naïve 2.
- The Mean Absolute Scaled Error (MASE), introduced by Hyndman and Koehler (2006).

The overall evaluation involves the calculation and comparison of over 55 billion forecast errors.

### 3.2 Simulation on Intermittent Data

We also considered the case of intermittent demand data, where two main factors were considered, namely average *inter-demand interval* (*IDI*) and squared *coefficient of variation* ($CV^2$) of the non-zero demands as it is dictated by the work of Syntetos et al. (2005). On top of that, we examined also the effect of the length of the series (*number of* available *observations*) that is quite ignored in the respective literature as usually in practice these series are short. Lastly, in line with the investigation on fast-moving data, we study the effects of *forecasting horizon*. So the basic research question may now be narrowed to:

> RQ2: *How four factors (inter-demand interval, coefficient of variation, the number of observations and the forecasting horizon) affect, if at all, the*

*forecasting accuracy of time series extrapolation methods on intermittent data*?

To generate simulated series for testing the above-mentioned research question, each of the first three factors (*IDI*, $CV^2$ and the *number of observations*) was varied around six levels (see Table 2) while 10,000 series were generated at each level (ceteris paribus). Given the number of different combinations ($216=6^3$) considered, we examine in total 2,160,000 simulated time series. For each series, we produce forecasts for the next 12 periods.

**Table 2:** Levels for the three factors considered in the intermittent demand data

| **Components** | **Level** | | | | | |
|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** |
| IDI | 1.00 | 1.16 | 1.32 | 1.60 | 2.00 | 4.00 |
| $CV^2$ | 0.00 | 0.25 | 0.49 | 0.75 | 1.00 | 2.00 |
| Number of observations | 24 | 36 | 48 | 60 | 84 | 108 |

The procedure followed to generate the intermittent demand data is given below. For a selected level of *number of observations* (*l*) and level of *IDI*, we generate a vector which specifies the occurrence of non-zero demands as a Bernoulli distribution (Croston, 1972; Syntetos and Boylan, 2001), where $p=1/IDI$. The output of this step is a binary vector of length *l*. For the demand sizes we use randomly generated numbers following a negative binomial distribution (see Syntetos et al., 2011) and increase them by one (1) as not to generate zero demands. The number of successful trials (*n*) and the success probability (*p*) can be easily derived as:

$$n = \frac{\mu p}{(1-p)} \qquad p = \frac{\mu}{c_v^2(\mu+1)^2} \tag{7}$$

where $c_v>0$ is the required *coefficient of variation* (square root of $CV^2$) and $\mu$ is the mean, selected randomly in [10, 50]. Letting the *l* generated values of the negative binomial distribution be the vector *d*, the required non-zero demand sizes can be derived as $D=d+1$. Finally $X_t$ is calculated as $X_t = I_t * D_t$ for t=1,…, *l*. If $c_v=0$, then $d=round(\mu)$ for every *t*, where *round* denotes the rounding function.

The forecasting methods used in this simulation are (for more details on these methods, see appendix A):

**Naïve**, Simple Moving Averages (**SMA**) of length 4, 8 and 12, **SES** with a prefixed smoothing parameter equal to 0.1 (SES(0.1)), **SES** with optimized smoothing parameter (SES(auto)), **Croston's method**, Syntetos-Boylan Approximation (**SBA**) and Teunter-Syntetos-Babai method (**TSB**).

For the methods designed specifically for intermittent demand (Croston, SBA, TSB), small values for the smoothing parameters were applied, as suggested by the literature (Syntetos and Boylan, 2005). In more detail, the smoothing parameter for estimating both demands and intervals in Croston's method and SBA and demands in TSB method is set equal to 0.1. Moreover, the smoothing parameter for estimating the probability of the occurrence of a non-zero demand in TSB method is set equal to 0.02.

The forecasting accuracy was measured by comparing the 12 hold out data points with the 12 point forecasts of each method. Due to the presence of zero demands, the use of percentage errors is not appropriate. We, therefore, use a scaled version of the Mean Absolute Error (sMAE), where the scaling is performed through dividing with the in-sample mean demand. The point forecast error (scaled absolute errors or sAE) for the *h*-step-ahead forecast can be calculated as follows:

$$sAE_{i,h} = \frac{\left| X_{N+h} - F_h \right|}{\frac{1}{N}\sum_{t=1}^{N} X_t} \tag{8}$$

where $X$ is the vector of observations generated previously, $F$ is the vector of point forecasts and $N$ is the length of the in-sample. The same is simply derived by averaging across horizons and series.

## 4. The Results: measuring the Influence of the factors

In this section we present results from simulated data for fast-moving and intermittent series, as well as for real fast-moving data from the M3 competition.

### 4.1 Simulations on fast-moving data

The accuracy results for the entire dataset are summarised in Table 3.

**Table 3:** Average sMAPE (**bold**) for all simulated fast-moving data [and MASE (*italics*), Percentage Better (versus Naïve 2) for 1 to 18 (underlined)]

| **Methods** | **sMAPE** (per *Forecasting Horizon*) | | | | | | | **M3 - Average sMAPE 1-18** | **Ratio: sMAPE M3/Simulated** | ***Average MASE*** | **Average Percentage Better** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **1** | **6** | **12** | **18** | **1-6** | **1-12** | **1-18** | | | *1-18* | 1-18 |
| **Naïve 1** | 5.89 | 15.05 | 6.60 | 16.58 | 9.09 | 10.88 | **10.71** | | | *1.39* | 15.20% |
| **Naïve 2** | 5.36 | 6.45 | 6.60 | 7.96 | 5.85 | 6.31 | **6.67** | 16.91 | 2.54 | *1.09* | - |
| **Single** | 4.61 | 5.43 | 6.00 | 7.00 | 5.01 | 5.43 | **5.81** | 15.32 | 2.64 | *1.01* | 55.70% |
| **Holt** | 4.67 | 5.58 | 6.53 | 7.70 | 5.13 | 5.66 | **6.20** | 15.36 | 2.48 | *0.98* | 55.50% |
| **Damped** | 4.61 | 5.42 | 5.96 | 6.93 | 5.01 | 5.42 | **5.78** | 14.59 | 2.52 | *0.95* | 58.50% |
| **Holt-Winters** | 4.69 | 5.76 | 6.98 | 8.57 | 5.25 | 5.89 | **6.59** | 15.44 | 2.34 | *1.05* | 52.00% |
| **Theta** | 4.62 | 5.32 | 5.89 | 6.65 | 4.98 | 5.35 | **5.69** | 13.85 | 2.43 | *0.94* | 60.20% |
| **Linear Trend** | 5.63 | 6.09 | 6.77 | 7.39 | 5.88 | 6.19 | **6.53** | 19.78 | 3.03 | *1.13* | 51.20% |
| **Autobox** | 5.40 | 6.25 | 6.53 | 8.04 | 5.66 | 6.16 | **6.55** | 15.83 | 2.42 | *1.04* | 49.90% |
| **Forecast Pro** | 4.62 | 5.42 | 6.13 | 7.14 | 5.03 | 5.46 | **5.88** | 13.86 | 2.36 | *0.93* | 56.70% |
| **SD** | 4.61 | 5.42 | 5.96 | 6.95 | 5.00 | 5.42 | **5.79** | | | *0.97* | 58.80% |
| **SHD** | 4.61 | 5.39 | 5.96 | 6.88 | 5.00 | 5.41 | **5.78** | | | *0.95* | 59.60% |
| **ST** | 4.61 | 5.36 | 5.91 | 6.77 | 4.98 | 5.37 | **5.72** | | | *0.96* | 60.70% |
| **SDT** | 4.60 | 5.37 | 5.91 | 6.80 | 4.98 | 5.38 | **5.73** | | | *0.96* | 60.70% |
| **SHDT** | 4.61 | 5.36 | 5.92 | 6.78 | 4.99 | 5.38 | **5.73** | | | *0.94* | 60.00% |

We note:

- The very good performance of the five combinations, a finding consistent with the conclusions from the M-Competitions.
- The strong similarities in the performance of methods (in terms of *sMAPEs)* with comparison to the ones in the M3-Competition, as depicted by the close values of the respective ratios.
- There are three single methods that consistently perform better: Single Exponential Smoothing[2], Damped Exponential Smoothing and the Theta method.
- The quite similar average results for *MASE* and *Percentage Better* as presented in the last two columns of Table 3.

[2] Also referred in the literate as 'Simple' Exponential Smoothing, or just abbreviated as SES and it is equivalent to an ARIMA(0,1,1) without constant model.

***Regressions***

Table 4 shows the results of the regression analysis. It lists the standardized *beta* coefficients for each factor, the corresponding *t-tests* as well as the overall $R^2$ and standard errors. Table 4 suggests that for all variables and methods the regression coefficients are statistically significant (at 0.01) with Naïve 1 being the only exception. Moreover the $R^2$ values range from 0.850 to 0.932, indicating a very good fits as expected from such a rich dataset.

**[Insert Table 4 about here]**

A positive *beta* coefficient means less accuracy whilst a negative one means improvement. Furthermore, the bigger the absolute value of the coefficient the greater the deterioration or improvement in forecasting performance. The signs of most regression coefficients are positive (*seasonality*, *cycle*, *randomness* and the *forecasting horizon*) and this means that when these following factors increase, the forecasting accuracy for all methods and combinations decreases.

Notable exceptions are the negative *betas* for: a) the *number of observations* factor for all methods with the exception Naïve 1, thus when the length of the series increases the accuracy increases as well even if it is a marginal improvement, and b) the *trend* factor for Holt, Holt-Winters and Linear Trend methods as these methods capture the trend in the data (see Table 4, column 4), and therefore marginally improve accuracy.

*Randomness* is the variable that most affects forecasting accuracy. Moreover, the values of these coefficients for the majority of methods are similar (ranging from 0.823 to 0.878); this means that the accuracy of all methods rapidly decreases as the randomness in the data increases, but also that practically all methods are equally capable of dealing with increasing levels randomness. The variable with the least influence is *trend*, in particular for the Holt, Holt-Winters and Linear Trend methods as the values of the corresponding regression coefficients are small. *Cycle* and the *Forecasting Horizon* variables appear to have less influence in terms of the extent to which they affect forecasting accuracy. Furthermore, the *seasonal* fluctuations in the data are captured in a similar way by practically all methods, as their regression coefficients are small, ranging from 0.026 to 0.063 (the exceptions are Naïve 1 and one of the commercial packages).

### 4.2 Application in real data

One way to apply the findings of the previous section in real data is as follows: when decomposing a time series we can estimate the levels of *seasonality*, *cycle*, *trend*, and *randomness* while we also do know exactly the *number of* available *observations* of the series and we decide on the *forecasting horizon*. This information allows us to estimate the percentage error for each forecasting horizon by utilizing the corresponding regression equation for each method and combination of methods shown in Table 4. Consequently we can identify the method/combination with the smallest error as well rank all methods according to their respective errors. We can, therefore, select for each series and forecasting horizon the method or combination with the minimum expected error.

Often the difference between the best method/combination identified and the second, third and fourth are small and to account for this we consider the simple combination of these methods/combinations for which a specific criterion - hereafter called *Threshold Ratio* - is more than a pre-defined value. The *Threshold Ratio* may be calculated by dividing the smallest expected sAPE of the 'optimal' method/combination for a given forecasting horizon with the expected sAPE of the second, third and fourth method/combination respectively. The maximum number of models to be combined could be limited and in this example we use the ad-hoc value of 6, as combinations of large pools of methods is not regarded as beneficial (Fildes and Petropoulos, 2013).

We apply this model selection and combination approach to the monthly series of the M3-Competition (1,428 time series, Makridakis and Hibon, 2000). The results are presented in Table 5 illustrating that the proposed model selection protocol, based on the regression estimation in Table 4, results in improved forecasting performance of 9.9% and 5.4% compared to SES and Damped respectively. For low values of the threshold so as to allow more models to be included in the simple combination, the performance gets even better than that of Theta model (Figure 1).

**Table 5.** Forecasting performance of the model selection protocol (sMAPE %)

| Models | | Forecasting Horizon | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Model Selection | Threshold | **1** | **2** | **3** | **1 to 4** | **1 to 8** | **1 to 12** | **1 to 18** |
| | 0.4 | 11.7 | 10.7 | 12.2 | 11.93 | 12.10 | 12.49 | 13.80 |
| | 0.5 | 11.8 | 10.8 | 12.3 | 12.06 | 12.19 | 12.57 | 13.86 |
| | 0.6 | 11.9 | 10.9 | 12.5 | 12.14 | 12.32 | 12.75 | 14.08 |
| | 0.7 | 12 | 11.1 | 12.6 | 12.35 | 12.54 | 13.05 | 14.48 |
| | 0.8 | 12.1 | 11.2 | 12.7 | 12.45 | 12.67 | 13.30 | 14.86 |
| | 0.9 | 12.3 | 11.1 | 13 | 12.59 | 12.89 | 13.62 | 15.30 |
| | 1 | 12.3 | 11.3 | 13 | 12.75 | 13.14 | 13.96 | 15.90 |
| SES | | 13 | 12.1 | 14 | 13.53 | 13.60 | 13.83 | 15.32 |
| Damped | | 11.9 | 11.4 | 13 | 12.63 | 12.85 | 13.10 | 14.59 |
| Theta | | 11.2 | 10.7 | 11.8 | 11.54 | 12.13 | 12.50 | 13.85 |

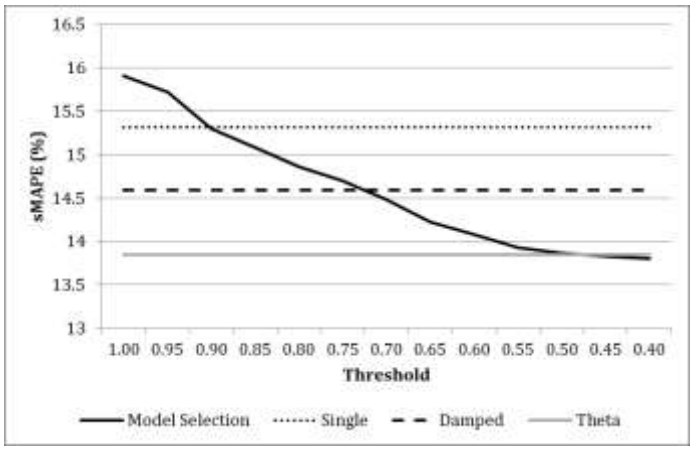


**Figure 1.** Forecasting performance of the model selection protocol on real data (M3-monthly data)

## 4.3 Simulations on Intermittent Data

The *sMAE* results for the entire dataset are shown in Table 6.

**Table 6.** Average sMAE for all simulated intermittent demand data

| **Methods** | **Forecasting Horizon** | | | | | **Average** | | |
|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **6** | **12** | **1 to 3** | **1 to 6** | **1 to 12** |
| Naïve | 1.213 | 1.211 | 1.212 | 1.213 | 1.210 | 1.212 | 1.213 | 1.212 |
| SMA(4) | 1.070 | 1.068 | 1.068 | 1.069 | 1.068 | 1.069 | 1.069 | 1.068 |
| SMA(8) | 1.037 | 1.036 | 1.036 | 1.036 | 1.035 | 1.036 | 1.036 | 1.036 |
| SMA(12) | 1.026 | 1.024 | 1.024 | 1.025 | 1.024 | 1.025 | 1.025 | 1.024 |
| SES(0.1) | 1.017 | 1.015 | 1.015 | 1.016 | 1.015 | 1.016 | 1.016 | 1.015 |
| SES(auto) | 1.016 | 1.015 | 1.014 | 1.015 | 1.014 | 1.015 | 1.015 | 1.015 |
| Croston | 1.026 | 1.024 | 1.024 | 1.024 | 1.023 | 1.024 | 1.024 | 1.024 |
| SBA | 1.015 | 1.013 | 1.013 | 1.014 | 1.013 | 1.014 | 1.014 | 1.013 |
| TSB | 0.990 | 0.988 | 0.988 | 0.988 | 0.987 | 0.988 | 0.988 | 0.988 |

There are a few interesting findings in Table 6:

- TSB performs slightly better than the other methods for all respective horizons.
- The improvement in forecasting accuracy as the length of the Simple Moving Average increases.
- *Forecasting horizon* does not affect accuracy.

***Regressions***

In Table 7 we present the results of the multiple regression analysis where the standardized *beta* coefficients and their corresponding values of *t-test* are listed for each factor. Moreover, the overall goodness of fit ($R^2$) and the standard error estimates are provided. The reported $R^2$ values indicate very good fit for all equations as in the case of fast-moving data.

**Table 7:** Multiple regression analysis for intermittent data: Dependent variable is sMAE (Standardized Coefficients)

| **Methods** | **IDI** | | **$CV^2$** | | **Number of Observations** | | **Forecasting Horizon** | | **$R^2$** | **Std Error** |
|---|---|---|---|---|---|---|---|---|---|---|
| | **$b_1$** | **$t_1$** | **$b_2$** | **$t_2$** | **$b_3$** | **$t_3$** | **$b_4$** | **$t_4$** | | |
| Naïve | 0.722 | 90.4 | 0.557 | 69.7 | -0.061 | -7.6 | -0.001 | -0.2 | 0.835 | 0.193 |
| SMA(4) | 0.797 | 120.8 | 0.499 | 75.6 | -0.062 | -9.3 | -0.001 | -0.2 | 0.887 | 0.153 |
| SMA(8) | 0.814 | 126.8 | 0.476 | 74.2 | -0.063 | -9.8 | -0.001 | -0.2 | 0.893 | 0.147 |
| SMA(12) | 0.820 | 129.1 | 0.468 | 73.6 | -0.063 | -9.9 | -0.001 | -0.2 | 0.896 | 0.145 |
| SES(0.1) | 0.825 | 130.7 | 0.460 | 72.9 | -0.064 | -10.2 | -0.001 | -0.2 | 0.897 | 0.143 |
| SES(auto) | 0.827 | 132.9 | 0.458 | 73.5 | -0.080 | -12.8 | -0.001 | -0.2 | 0.900 | 0.143 |
| Croston | 0.835 | 135.0 | 0.445 | 71.9 | -0.077 | -12.5 | -0.001 | -0.2 | 0.901 | 0.146 |
| SBA | 0.841 | 137.7 | 0.437 | 71.5 | -0.080 | -13.0 | -0.001 | -0.2 | 0.904 | 0.140 |
| TSB | 0.797 | 114.8 | 0.488 | 70.2 | -0.040 | -5.8 | -0.001 | -0.2 | 0.875 | 0.144 |

Forecasting performance for all methods is heavily affected by the increase of intermittence of the data (*IDI*) as well as the *coefficient of variation*. As the values of these two factor increase, the respective accuracy of all models decreases. Even if the differences are very small, Croston and SBA are the two methods most affected by the increase of the average *IDI*, despite the fact that both methods are specifically designed for intermittent demand data. On the other hand, Naive is the method which is least affected by the *IDI*, as increase of intermittency results in spot-on forecasts (for zero demand). The exact opposite is true for the *coefficient of variation*. SBA and

Croston are the two methods with the lowest effect of on their forecasting accuracy. On the contrary, Naïve, SMA(4) and TSB are the three methods affected the most by variability in demand, according to the standardized *beta* coefficients.

The *number of* available *observations* has a very small positive impact on the forecasting accuracy. Finally, *forecasting horizon* has practically no impact on the predictive power of the alternative methods (non-statistically significant beta coefficients).

## 5. Discussion

For regular/fast-moving data, three relatively simple methods, Single and Damped exponential smoothing as well as the Theta method, demonstrated the best performance for all the three accuracy measures used. In between them, Theta performed better for longer forecasting horizons, the reason being its ability to more accurately predict the trend in the data, whilst Single and Damped exponential smoothing were more accurate for shorter forecasting horizons. Moreover, the five combinations of forecasting methods performed well, often exceeding the accuracy of the individual methods being combined, whilst also reducing the variance of forecasting errors. The above three findings are consistent and corroborate to the findings of the literature on empirical forecasting competitions.

The regression coefficients as presented in Table 4 allow us to identify which factors affecting methods' accuracy as well as measure the extent of such influences. *Cycle* and *randomness* have the biggest effect on the forecasting accuracy of all methods. It is interesting that Naïve 1 exhibits the best performance in respect to these two factors (it has the smallest regression coefficients of all methods), which could be attributed to the random walk nature of many of the simulated data series. *Seasonality* is well captured by all methods with the expected exception of Naïve 1. The similarity of the regression coefficients for this specific factor is due to the fact that the seasonal indexes, with the exception of Holt-Winters, were estimated using the classical decomposition approach. As expected, *trend* is best captured by Holt, Holt-Winters and Linear Regression, whilst Single, Naïve and Naïve 2 are unable to do so.

The regression coefficients of the *number of observations* factor, although negative, have a minimal effect on improving forecasting accuracy. The *forecasting horizon* factor coefficient indicates that accuracy decreases the longer the forecasting

horizon, providing a considerable advantage to Theta whose *beta* coefficient is smaller than that of all other methods except for that for Linear Trend. This later one however gets worse for longer horizons because of not being able to accurate capture and predict cyclical changes.

The poor performance, in contrast to other empirical studies in the literature, of the two commercial packages may be the result of three factors. Firstly, Autobox and Forecast Pro were tested only in 300 time series for each of the 7,776 cases (i.e. for a total of 2,332,800 time series versus 77,760,000 for all the other methods) due to time and computational constraints. Secondly, Tom Reilly (personal communications) pointed out that the poor performance of Autobox could be explained by the deterministic character of the models used to generate the data however the authors believe that this could only be partially true as the simulated data were derived from ranges and values from the real M3 data, where there as well Autobox was not one of the top-performing methods. So maybe the very range of values could be blamed but not the deterministic nature. Lastly, the commercial packages were used in fully automatic mode, with no supervision at all in this study and thus their parameters were not manually adjusted for the features of the generated data. We believe that this latter argument provides the best explanation for the software packages not being en par with the benchmarks.

We believe that through the regression analysis we can isolate the influence of each factor on forecasting accuracy, in a ceteris paribus fashion. This means that we can determine the most/least important factors influencing forecasting accuracy. *Seasonality*, for instance, seems to be able to be captured through the classical decomposition process. *Trend*, on the other hand, seems to be a much bigger challenge, as random and cyclical changes make the identification of a robust trend very difficult and, for this reason, Single and even Naïve 2 perform so well. Furthermore, additional historical information in the form of more *observations* and lengthier series improves accuracy but to a small extent, whilst *cyclical* fluctuations are found to play the most important role in forecasting accuracy. Furthermore, *randomness* also plays a significant role, but this is to be expected to a large extent and hard to deal with. Finally, combining forecasts appears to improve forecasting accuracy in the majority of cases whilst also reducing the variance of forecasting errors, highlighting that it is almost impossible to identify one single optimal model based on the data features.

For intermittent data there were no big surprises given the lot of attention in the recent year in the respective literature and the plethora of empirical investigations. The negative effects on forecasting accuracy from marginal increases of *IDI* or $CV^2$ were more or less expected. However, the performance of TSB is remarkable and an interesting finding given the extended simulation exercise. A few more remarkable findings were surfaced: first and foremost further research should shed light why the *length* of the series has so small impact and even more why the forecasting *horizon* is not influencing at all the forecasting accuracy.

## 6. Practical Implications for Decision Makers

One of the most important challenges of any academic article is how the results of the proposed research can be translated into practical recommendations for decision makers. To that end, we are in the comfortable position to claim to be informing real-life professionals on the appropriate uses of 'Horses for Courses' in demand forecasting. While the fundamental question still remains: "what is the best method for *my data*?", we believe now, and through the illustration of the use of the protocol in real data with improved forecasting performance (see section 4.2), that we are in a position to partially answer this question.

In essence, through this study we are providing guidance to practitioners on which are the most appropriate forecasting methods, given the specific data features they are facing. In order to accommodate this, we provide a graphical representation of the main results of our study (Figure 2). The numerical standardized *beta* coefficients of Table 4 have been converted into an eleven-scale (-5 to +5), representing the effect of each factor (*seasonality, trend, cycle, randomness, number of observations* and *forecasting horizon*) upon forecasting accuracy for the most popular forecasting methods among practitioners.

So, how this Figure 2 should be used from practitioners? First they need to decompose their time series so as to find out its *seasonality, cycle, trend* and *randomness* (whilst also *number of observations* and the *forecasting horizon* are known in advance). This information will allow them through Figure 2 to select the most appropriate methods amongst the ten presented in this study, the methods that are expected to end up with the lowest out-of-sample (*sMAPE)* and thus the best forecasting accuracy.

| Method | Seasonality | Trend | Cycle | Randomness | Number of observations | Forecasting Horizon |
|---|---|---|---|---|---|---|
| Naïve | ✖✖✖ | ✖ | ✖ | ✖✖ | | ✖ |
| Naïve 2 | ✖ | ✖ | ✖✖ | ✖✖✖✖✖ | ✓ | ✖ |
| Single | ✖ | ✖ | ✖✖ | ✖✖✖✖✖ | ✓ | ✖ |
| Holt | ✖ | | ✖✖✖ | ✖✖✖✖✖ | ✓ | ✖✖ |
| Damped | ✖ | ✖ | ✖✖ | ✖✖✖✖✖ | ✓ | ✖ |
| Holt-Winters | ✖ | | ✖✖✖ | ✖✖✖✖✖ | ✓ | ✖✖ |
| Theta | ✖ | ✖ | ✖✖ | ✖✖✖✖✖ | ✓ | ✖ |
| Linear Trend | ✖ | ✓ | ✖✖✖ | ✖✖✖✖ | ✖ | ✖ |
| Autobox | ✖ | ✖ | ✖✖ | ✖✖✖✖✖ | ✓ | ✖ |
| Forecast Pro | ✖ | ✖ | ✖✖ | ✖✖✖✖✖ | ✓ | ✖✖ |

**Figure 2:** The **Method Selection Protocol** for **fast-moving** data.

**(**✖=decreasing accuracy where ✖=*0.5*✖, ✓=increasing accuracy where ✓=*0.5*✓**)**

For example, following the **Method Selection Protocol** for regular/fast-moving data in Figure 2, if *seasonality* and *trend* are the only features present in the data, then Holt, Holt-Winters and Linear Trend should be selected, with Linear Trend being the best approach. When *trended* data with many *observations* are to be extrapolated, Holt seems the most obvious choice. In the case of a strong presence of *cycle*, *randomness* or both, Naïve is by far the best option. If *cycle* is present as well as *seasonality*, then Naïve 2 is the method to select. On the other hand, if *cycle* is the dominant feature, while we are dealing with time series with many data points, then Linear Trend should be avoided. In the case of large forecasting horizons, Holt-Winters, Holt and Forecast Pro are not considered as good options. On the contrary, Naïve and Theta should be selected. Concluding, this specific protocol could be used from practitioners as a broad 'rule of thumb' for method selection given their specific data.

When it seems from the protocol that many methods could be used for a specific data series then a simple combination of those methods could well be used instead. The benefit from the aforementioned combination is twofold as the variance of the forecasting errors is also reduced.

The respective Method Selection Protocol for Intermittent data is illustrated in Figure 3, which is practically the graphical equivalent of Table 7.

| Method | IDI | $CV^2$ | Number of observations | Forecasting Horizon |
|---|---|---|---|---|
| Naïve | ✖✖✖✖ | ✖✖✖✖ | | - |
| SMA(4) | ✖✖✖✖✖ | ✖✖✖ | | - |
| SMA(8) | ✖✖✖✖✖ | ✖✖✖ | | - |
| SMA(12) | ✖✖✖✖✖ | ✖✖✖ | | - |
| SES(0.1) | ✖✖✖✖✖ | ✖✖✖ | | - |
| SES(auto) | ✖✖✖✖✖ | ✖✖✖ | ✓ | - |
| Croston | ✖✖✖✖✖ | ✖✖✖ | ✓ | - |
| SBA | ✖✖✖✖✖ | ✖✖✖ | ✓ | - |
| TSB | ✖✖✖✖✖ | ✖✖✖ | | - |

**Figure 3:** The **Method Selection Protocol** for **intermittent** data.
**(**✖=decreasing accuracy where ✖=*0.5*✖, ✓=increasing accuracy where ✓=*0.5*✓**)**

By and large for high values of *IDI* forecasters are prompted to use either TSB or SMA (Naïve is not suggested, as this method will result in under-stocks), while for high values of $CV^2$ SES or Croston/SBA should be used. In the unlikely event of having many *observations*, then yet again SES or Croston/SBA should be used.

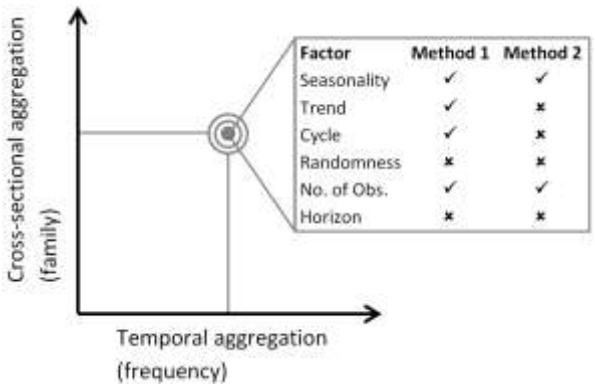


**Figure 4:** Given a specific level of **temporal** and **cross-sectional aggregation** level , the formed time series will have some characteristics based on which the **Method Selection Protocol** can be applied.

It needs to be emphasized that the **Method Selection Protocol** can be applied for any level of temporal (**frequency)** and hierarchical/family **cross-sectional aggregation**. Forecasters might select a specific frequency for either reasons of plain convenience or scientifically driven from the need to improve forecasting accuracy (see the ADIDA framework, Nikolopoulos et al. 2011), and in a similar fashion a

specific level of cross-sectional aggregation. As a result, a specific time series will be constructed; this time series will have specific features and, based on these features, the Model Selection Protocol could well be applied as illustrated in Figure 4.

## 7. Conclusion and Future Research

One of the biggest challenges that forecasters and practitioners are constantly facing is the selection of the most appropriate method for a specific data set or even for a single time series. The quest for forecasting approaches tailored to specific types of data, led researchers to the development of method selection protocols, usually based on a large number of rules necessitating advanced data analysis.

In this study we assume that seven time series features *(seasonality, trend, cycle, randomness, number of observations/length, IDI, $CV^2$*) and one strategic decision (*forecasting horizon*) are the dominant determinants of forecasting accuracy. Through two extensive simulations with almost 80 million regular/fast-moving and intermittent time series, we study the accuracy of the most popular forecasting methods and measure the factors that affect their respective performance. Fourteen forecasting methods plus five combinations of them are employed in order to predict twelve to eighteen periods ahead and respectively measure the out-of-sample forecasting accuracy using four fit-for-purpose metrics (*sMAPE, Percentage Better, MASE* and *sMAE*). Consequently, regressions analysis was performed so as to determine the sign and the amplitude of the effect of each of the factors on the performance of the evaluated methods. Our main findings conclude that in terms of the achieved forecasting accuracy:

For regular/fast-moving data – where demand occurs each and every period:

- *Cycle* and *randomness* have the biggest (negative) effect on forecasting accuracy.
- The longer the *forecasting horizon*, the more accuracy decreases.

For intermittent data:

- Increasing *IDI* has the biggest (negative) effect in accuracy.
- Increasing $CV^2$ has also a negative effect.

For all types of data

- Increasing the length of the series has a small positive effect in accuracy.

One of the main practical contributions of the current work is that it translates the statistical findings into a *graphical* **Method Selection Protocol** that enables decision makers and practitioners to select the most appropriate forecasting method for their own data.

As far as the future of similar investigations is concerned, it is important to expand the pool of both methods and factors/data features being considered. Paths for future research could also include:

- Evaluating density forecasts rather than point forecasts in similar simulation setups.
- Introducing temporary and permanent structural changes in the level and trend of the data and determine their impact on accuracy.
- Running sliding simulations in the forecast evaluations (Tashman, 2000).

*Epilogue*: as the quest for '*Horses for Courses*' in demand forecasting is a long standing issue in the forecasting literature, we hope that we have shed some light towards that direction…


## Acknowledgements

We would like the thank: the handling editor Professor Immanuel Bomze for his insightful comments, three anonymous referees for their constructive comments, Professor Aris Syntetos and Professor Dimitrios Thomakos for their comments on earlier versions of this paper, Tom Reilly and Eric Stellwagen from Autobox and Forecast Pro respectively for providing us with evaluation versions of their commercial forecasting software.


## References


Adam, E. E. (1973). Individual item forecasting model evaluation. *Decision Sciences*, *4*, 458-470.

Adya, M., Armstrong, J. S., Collopy, F., & Kennedy, M. (2000). An application of rule-based forecasting to a situation lacking domain knowledge. *International Journal of Forecasting*, *16*, 477-484.

Adya, M., Collopy, F., Armstrong, J. S., & Kennedy, M., (2001). Automatic identification of time series features for rule-based forecasting. *International Journal of Forecasting*, *17*, 143-157.

Assimakopoulos, V., & Nikolopoulos, K. (2000). The theta model: a decomposition approach to forecasting. *International Journal of Forecasting, 16,* 521-530.

Athanasopoulos, G., Ahmed, R. A., & Hyndman, R. J. (2009). Hierarchical forecasts for Australian domestic tourism. *International Journal of Forecasting, 25*, 146-166.

Athanasopoulos, G., Hyndman, R. J., Song, H., & Wu, D. C. (2011). The tourism forecasting competition. *International Journal of Forecasting, 27*, 822-844.

Babai, M.Z., Ali, M., & 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.

Bacchetti, A., & Saccani, N. (2012). Spare parts classification and demand forecasting for stock control: Investigating the gap between research and practice. *Omega: The International Journal of Management Science*, *40*, 722-737.

Billah, B., King, M. L., Snyder, R. D., & Koehler, A. B. (2006). Exponential smoothing model selection for forecasting. *International Journal of Forecasting 22*, 239-247.

Box, G., & Jenkins, G. (1970). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.

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

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

Broyles, J. R. Cochran, J.K., & Montgomery, D. C. (2010). A statistical Markov chain approximation of transient hospital inpatient inventory. *European Journal of Operational Research, 207*, 1645–1657.

Cang, S., & Yu, H. (2014). A combination selection algorithm on forecasting. *European Journal of Operational Research, 234*, 127–139.

Cao, Q., Ewing, B. T., & Thompson, M. A.(2012). Forecasting wind speed with recurrent neural networks. *European Journal of Operational Research*, *221*, 148–154.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography (with discussion). *International Journal of Forecasting, 5*, 559-583.

Collopy, F., & Armstrong, J. S. (1992). Rule-Based Forecasting: development and validation of an expert systems approach to combining time series extrapolations. *Management Science*, *38*, 1394-1414.

Cooper, R. L. (1972). The predictive performance of quarterly econometric models of the United States. In *Econometric Models of Cyclical Behavior, Studies in Income and Wealth*, *11*, 813-925, Hickman BG. (Ed.). New York: Columbia University Press.

Crone, S. F., Hibon, M., & Nikolopoulos, K. (2011). Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction. *International Journal of Forecasting*, *27*, 635-660.

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

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*, 571-582.

Fildes, R. (1989). Evaluation of aggregate and individual forecast method selection rules. *Management Science*, *39*, 1056-1065.

Fildes R., & Makridakis S. (1995). The impact of empirical accuracy studies on time series analysis and forecasting. *International Statistical Review*, *63*, 289-308.

Fildes R., & Petropoulos F. (2013). An evaluation of simple versus complex selection rules for forecasting many time series. *Journal of Business Research*, forthcoming.
Franses, P. H., & van Dijk, D. (2005). The forecasting performance of various models for seasonality and nonlinearity for quarterly industrial production. *International Journal of Forecasting*, *21*, 87-102.
Gardner, Jr. E. S., & McKenzie, E. (1985). Forecasting trends in time series. *Management Science*, *31*, 1237-1246.
Gardner, Jr. E. S., & McKenzie, E. (1988). Model identification in exponential smoothing. *Journal of the Operational Research Society*, *3*, 863-867.
Gardner, Jr. E. S. (2006). Exponential smoothing: The state of the art-Part II. *International Journal of Forecasting*, *22*, 637-666.
Gardner, Jr. E. S., & Diaz-Saiz, J. (2008). Exponential smoothing in the telecommunications data. *International Journal of Forecasting, 24*, 170-174.
Gil-Alana, L. A., Cunado, J., & De Garcia, F. P. (2008). Tourism in the Canary Islands: forecasting using several seasonal time series models. *Journal of Forecasting*, *27*, 621-636.
Groff, G. K. (1973). Empirical comparison of models for short-range forecasting. *Management Science*, *20*, 22-31.
Hibon, M., & Evgeniou, T. (2005). To combine or not to combine: selecting among forecasts and their combinations. *International Journal of Forecasting, 21*, 15-24.
Hogarth, R. (1987). *Judgement and Choice* (second edition). New York: Wiley.
Holt, C. C. (1957). *Forecasting seasonals and trends by exponentially weighted averages*. O. N. R. Memorandum 52/1957. Pittsburgh: Carnegie Institute of Technology. Reprinted with discussion in 2004. *International Journal of Forecasting*, *20*, 5-13.
Hyndman, R. J., & Billah, B. (2003). Unmasking the Theta method. *International Journal of Forecasting*, *19*, 287-290.
Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*, 679-688.
Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, *18*, 439-454.
Jose, V. R. R., & Winkler, R. L. (2008). Simple robust averages of forecasts: Some empirical results. *International Journal of Forecasting*, *24*, 163-169.
Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*, 7-251.
Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.
Kirby, R. M. (1966). A comparison of short and medium range statistical forecasting methods. *Management Science*, *13*, 202-210.
Kolassa, S. (2011). Combining exponential smoothing forecasts using Akaike weights. *International Journal of Forecasting*, *27*, 238-251.
Kostenko, A. V., & Hyndman, R. J., (2006). A note on the categorization of demand patterns. *Journal of the Operational Research Society, 57*, 1256–1257.
Kourentzes N., Petropoulos F., & Trapero J. R. (2014). Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting*, *30*, 291-302
Krampf, R. F. (1972). *The turning point problem in smoothing models*. PhD Thesis. University of Cincinnati: Ohio.
Levine, A. H. (1967). Forecasting techniques. *Management Accounting*, 86-95.

Makridakis, S., & Hibon, M. (1979). Accuracy of Forecasting: An Empirical Investigation. *Journal of the Royal Statistical Society*, Series A, *142*, 97-145.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., & Winkler, R. (1982). The accuracy of extrapolation (time series) methods: results of a forecasting competition. *Journal of Forecasting*, *1*, 111-153.

Makridakis, S., & Winkler, R. (1983). Average of Forecasts: Some Empirical Results. *Management Science*, *29*, 987-996.

Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., Mills, T., Ord, K., & Simmons, L. F. (1993). The M-2 Competition: a real-time judgmentally based forecasting study. *International Journal of Forecasting*, *9*, 5-23.

Makridakis, S., & Hibon, M. (2000). The M3-Competition: Results, conclusions and implications. *International Journal of Forecasting*, *16*, 451-476.

Meade, N., (2000). Evidence for the selection of forecasting methods. *Journal of Forecasting*, *19*, 515-535.

Meehl, P. (1954). *Clinical Versus Statistical Predictions: A Theoretical Analysis and a Review of the Evidence*. Minneapolis: Minneapolis University Press.

Meehl, P. E. (1986). Causes and effects of my disturbing little book. *Journal of Personality Assessment*, *50*, 370–375.

Miller, D., & Williams, D. (2003). Shrinkage estimators of time series seasonal factors and their effect on forecasting accuracy. *International Journal of Forecasting*, *19*, 669-684.

Naylor, T. H., Seaks, T. G. (1972). Box-Jenkins methods: an alternative to econometric models. *International Statistical Review*, *40*, 113-137.

Newbold, P., & Granger, C. W. J. (1974). Experience with forecasting univariate time series and the combination of forecasts (with Discussion). *Journal of the Royal Statistical Society*, Series A, *137*, 131-165.

Nikolopoulos, K., & Assimakopoulos, V. (2003). Theta Intelligent Forecasting Information System. *Industrial Management and Data Systems*, *103*, 711-726.

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

Nikolopoulos, K., Syntetos, A., Boylan, J., Petropoulos, F., & Assimakopoulos, V. (2011). ADIDA: An aggregate/disaggregate approach for intermittent demand forecasting. *Journal of the Operational Research Society*, *62*, 544-554.

Pegels, C. C. (1969). Exponential Forecasting: Some New Variations. *Management Science*, *15*, 311-315.

Rostami-Tabar, B., Babai, M. Z., Syntetos, A. A., & Ducq, Y. 2013. Demand forecasting by temporal aggregation. *Naval Research Logistics*, *60*, 479-498.

Shah, C. (1997). Model selection in univariate time series forecasting using discriminant analysis. *International Journal of Forecasting*, *13*, 489-500.

Slovic, P. (1972). Psychological study of human judgement: implications for investment decision making. *Journal of Finance*, *27*, 779-799.

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

Spithourakis,, G. P., Petropoulos, F., Nikolopoulos, K., & Assimakopoulos, V. (2013). A systemic view of the ADIDA framework. *IMA Journal of Management*

*Mathematics, first published online December 2, 2012*, doi:10.1093/imaman/dps031
Surowiecki, J. (2005). *The Wisdom of Crowds*. New York: Anchor Book.
Syntetos, A. A., & Boylan, J. E. (2001). On the bias of intermittent demand estimates. *International Journal of Production Economics*, *71*, 457-466.
Syntetos, A. A., & Boylan, J. E. (2005). The accuracy of intermittent demand estimates. *International Journal of Forecasting*, *21*, 303-314.
Syntetos, A. A., Boylan, J. E., & 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., Lengu, D., & Altay, N. (2011). *Distributional assumptions for parametric forecasting of intermittent demand*. In Altay N., & Litteral L. A. (Eds.), *Service Parts Management* (pp. 31-52). London: Springer.
Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting*, *16*, 437-450.
Taylor, J. W. (2008). Exponentially weighted information criteria for selecting among forecasting models. *International Journal of Forecasting*, *24*, 513-524.
Teunter, R. H., Syntetos A., & Babai Z. (2011). Intermittent demand: Linking forecasting to inventory obsolescence. *European Journal of Operational Research*, *214*, 606-615.
Thomakos, D. D., & Nikolopoulos, K. (2014). Fathoming the Theta Method for a Unit Root Process. *IMA Journal of Management Mathematics, 25*, 105-124.
Tversky, A., & Kahneman, D. (1982). Judgement under uncertainty: heuristics and biases. In *Judgement under uncertainty: heuristics and biases*, 3-20, Kahneman, D., Slovic, P., Tversky A. (Eds.). Cambridge University Press: Cambridge.
Winters, P. R. (1960). Forecasting Sales by Exponentially Weighted Moving Averages. *Management Science*, *6*, 324-342.

## Appendix A: Forecasting Methods

**Naïve.** This is the simplest forecasting approach. Point forecasts for all lead times are equal to the latest available actual observation (random walk).

**Naïve 2.** The point forecasts are derived as in the Naïve method, using the seasonally adjusted data. A deterministic multiplicative seasonality is assumed. Finally, point forecasts are reseasonalised using the appropriate seasonal indices.

**Exponential Smoothing.** The exponential smoothing methods are based on averaging (smoothing) past values of a time series in a decreasing (exponential) manner. Single Exponential Smoothing (*Single*) assumes no trend or seasonal patterns (Brown, 1956). Forecasts for Single can be produced via the following formula:

$$F_{t+1} = aX_t + (1-a)F_t \tag{A.1}$$

where $\alpha$ refer to the exponential smoothing parameter for the level.

Holt Exponential Smoothing (*Holt*) expands Single adding one additional parameter for smoothing the short-term trend (Holt, 1957). The point forecasts for Holt can be calculated via:

$$L_t = aX_t + (1-a)(L_{t-1} + T_{t-1}) \tag{A.2}$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \tag{A.3}$$

$$F_{t+m} = L_t + mT_t \tag{A.4}$$

where $\beta$ is the smoothing parameter for the trend, $L_t$ refers to the forecast of the level for period $t$ and $T_t$ is the forecast for the trend for period $t$.

Damped Exponential Smoothing (*Damped*) introduces a dampening factor ($\varphi$) that is multiplied on the trend component of Holt's method in order to give more control over the long-term extrapolation of the trend (Gardner and McKenzie, 1985). Forecasts for Damped can be calculated as:

$$L_t = aX_t + (1-a)(L_{t-1} + \phi T_{t-1}) \tag{A.5}$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)\phi T_{t-1} \tag{A.6}$$

$$F_{t+m} = L_t + \sum_{i=1}^{m} \phi^i T_t \tag{A.7}$$

Single, Holt and Damped are applied on non-seasonal or seasonally adjusted data, where the final forecasts are reseasonalised. However, their performance is contrasted with *Holt*-Winters (Winters, 1960) a method which includes stochastic multiplicative modelling of seasonality. Forecasts for Holt-Winters are derived as follows:

$$L_t = a\frac{X_t}{S_{t-s}} + (1-a)(L_{t-1} + T_{t-1}) \tag{A.8}$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \quad \text{(A.9)}$$

$$S_t = \gamma \frac{X_t}{L_t} + (1-\gamma)S_{t-s} \quad \text{(A.10)}$$

$$F_{t+m} = (L_t + mT_t)S_{t-s+m} \quad \text{(A.11)}$$

where $S_t$ is the estimate of the seasonal index for period $t$ and $\gamma$ is the seasonal smoothing factor.

**Theta Model.** The Theta model (Assimakopoulos and Nikolopoulos, 2000, Thomakos and Nikolopoulos, 2013) decomposes the seasonally adjusted series into two so-called 'Theta lines'. The first Theta-line is calculated as the time regression of the data, thus corresponds to the long-term trend of the data. This Theta-line is extrapolated as usual (linear regression line). The second Theta-line has double the curvatures of the seasonally adjusted data and is simply calculated as 2*X*-*LRL*, where *X* is the vector of the seasonally adjusted data and *LRL* are the values of the linear regression line. The second Theta-line is extrapolated using Single Exponential Smoothing. The point forecasts if the two Theta-lines are combined using equal weights. The final forecasts are reseasonalised.

**Linear Regression.** This method assumes a relationship between the values (dependent variable) and the timestamps of the respective periods (independent variable). The mathematical model of this relationship is the linear regression equation. We calculate this equation using ordinary least squares. In this research, Linear Regression is applied on the seasonally adjusted data, while final forecasts are reseasonalised.

**Forecast Pro.** Forecast Pro's Expert Selection is a routine implemented in the Forecast Pro commercial package (www.forecastpro.com). This method analyses each time series and performs individual model selection.

**Autobox.** A fully automated Box-Jenkins model building process is implemented in the commercial package Autobox (www.autobox.com). This includes model identification, estimation and diagnostic feedback loop.

**Simple Moving Averages.** The point forecasts are calculated as an unweighted average of the last *k* observations, as follows:

$$F_{t+1} = \frac{1}{k} \sum_{i=t-k+1}^{t} X_i \quad \text{(A.12)}$$

**Croston's method.** Croston (1972), suggested the decomposition of an intermittent demand series in the non-zero observed demands and the intervals (in time periods) between successive non-zero demands. The demands and the intervals are extrapolated separately, using simple exponential smoothing parameter ($\alpha$) with relatively low smoothing value. Updating for both demands and intervals performs only for non-zero demands. The ratio of the two forecasts will constitute the final forecast:

$$F_{t+1} = \frac{\hat{z}_{t+1}}{\hat{p}_{t+1}} \tag{A.13}$$

where $\hat{z}_{t+1}$ and $\hat{p}_{t+1}$ are the forecasts for the demands and the intervals respectively for period $t$+1.

**Syntetos & Boylan Approximation (SBA).** Syntetos and Boylan (2001) proved that Croston's method is positively biased. Subsequently, they proposed (Syntetos and Boylan 2005) that final forecasts should be multiplied by a debiasing factor, which depends on the value of the smoothing parameter. The forecasts for their proposed approximation (SBA) can be derived as follows:

$$F_{t+1} = \left(1 - \frac{a}{2}\right)\frac{\hat{z}_{t+1}}{\hat{p}_{t+1}} \tag{A.14}$$

**Teunter, Syntetos & Babai (TSB) Method:** Teunter, Syntetos and Babai (2011) considered an alternative decomposition approach. They proposed that the forecasts of the non-zero demands should be multiplied by the forecast of the probability that a non-zero demand will occur, thus:

$$F_{t+1} = \hat{z}_{t+1} \cdot \hat{\rho}_{t+1} \tag{A.15}$$

where $\hat{\rho}_{t+1}$ is the forecast of the probability observing a non-zero demand. While $z$ updates when a non-zero demand occurs, $\rho$ will be updated every period.

## Appendix B: M3-Competition

The M3-Competition (Makridakis and Hibon, 2000) is still to date the largest empirical forecasting competition. The study compared the forecasting performance of 26 different approaches (19 research teams/benchmarks and 7 forecasting packages) on 3,003 time series. The data covered a range of frequencies (yearly, quarterly, monthly, other) and various types of time series (micro, macro, industry, etc.). The results of the M3-Competition have referred to numerous publications, with its data being used very often for empirical studies. Theta model (Assimakopoulos and Nikolopoulos, 2000) achieved the best performance from the academic contestants while Forecast Pro (http://www.forecastpro.com/) topped the table of commercial forecasting packages.

## Appendix C: Replicability

With regards to the replicability of this study, readers can download a special built software from http://www.forlab.eu/forecasting-software entitled 'Horses*for*Courses Simulator'. Simulated time series as well as forecasts for all methods considered in this research can be reproduced using the provided software, with the only exception being for forecasts produced from Forecast Pro and Autobox as these are copyrighted commercial software packages.

**Table 4:** Multiple regression analysis: the dependent variable is the sMAPE (standardized coefficients)

| sMAPE | Seasonality | | Trend | | Cycle | | Randomness | | Number of Observations | | Forecasting Horizon | | $R^2$ | Std. Error of the Estimate |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $b_1$ | $t_1$ | $b_2$ | $t_2$ | $b_3$ | $t_3$ | $b_4$ | $t_4$ | $b_5$ | $t_5$ | $b_6$ | $t_6$ | | |
| **Naïve 1** | 0.569 | 279.9 | 0.035 | 17.1 | 0.070 | 34.7 | 0.292 | 143.7 | 0.002 | *0.8* | 0.088 | 43.4 | 0.422 | 6.731 |
| **Naïve 2** | 0.043 | 61.5 | 0.054 | 76.8 | 0.241 | 344.7 | 0.916 | 1309.8 | -0.040 | -56.9 | 0.166 | 236.9 | 0.932 | 1.067 |
| **Single** | 0.031 | 40.4 | 0.071 | 92.7 | 0.323 | 420.6 | 0.874 | 1139.7 | -0.037 | -48.7 | 0.206 | 268.0 | 0.918 | 0.936 |
| **Holt** | 0.050 | 65.5 | -0.002 | -2.2 | 0.408 | 532.8 | 0.824 | 1076.1 | -0.101 | -131.7 | 0.244 | 318.7 | 0.918 | 1.072 |
| **Damped** | 0.033 | 43.3 | 0.056 | 74.0 | 0.342 | 448.5 | 0.870 | 1142.1 | -0.044 | -57.8 | 0.195 | 256.2 | 0.919 | 0.947 |
| **Holt-Winters** | 0.044 | 59.6 | -0.003 | -4.2 | 0.396 | 540.4 | 0.823 | 1123.5 | -0.052 | -71.3 | 0.295 | 402.7 | 0.925 | 1.096 |
| **Theta** | 0.030 | 40.2 | 0.024 | 32.1 | 0.346 | 459.0 | 0.875 | 1162.1 | -0.042 | -55.1 | 0.178 | 235.7 | 0.921 | 0.939 |
| **Linear Trend** | 0.026 | 29.2 | -0.012 | -13.5 | 0.549 | 612.4 | 0.747 | 834.1 | 0.041 | 45.7 | 0.160 | 178.3 | 0.888 | 1.171 |
| **Autobox** | 0.140 | 134.9 | 0.071 | 68.3 | 0.383 | 370.2 | 0.799 | 771.6 | -0.029 | -28.1 | 0.197 | 189.9 | 0.850 | 1.384 |
| **Forecast Pro** | 0.063 | 86.2 | 0.033 | 45.9 | 0.378 | 518.3 | 0.854 | 1170.3 | -0.055 | -76.1 | 0.214 | 292.7 | 0.926 | 0.931 |
| **SD** | 0.031 | 41.1 | 0.062 | 81.8 | 0.332 | 438.8 | 0.874 | 1155.9 | -0.040 | -52.5 | 0.199 | 262.8 | 0.920 | 0.932 |
| **SHD** | 0.037 | 50.0 | 0.034 | 45.5 | 0.352 | 471.7 | 0.870 | 1166.0 | -0.051 | -68.8 | 0.192 | 257.9 | 0.922 | 0.938 |
| **ST** | 0.031 | 40.9 | 0.046 | 60.8 | 0.333 | 442.9 | 0.878 | 1165.6 | -0.037 | -48.5 | 0.187 | 248.8 | 0.921 | 0.925 |
| **SDT** | 0.031 | 41.1 | 0.048 | 64.0 | 0.335 | 447.4 | 0.877 | 1169.5 | -0.038 | -50.2 | 0.188 | 250.9 | 0.921 | 0.924 |
| **SHDT** | 0.035 | 46.5 | 0.031 | 42.1 | 0.348 | 467.1 | 0.873 | 1170.7 | -0.047 | -62.8 | 0.186 | 249.3 | 0.922 | 0.933 |