Expert Systems With Applications 129 (2019) 10–26

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A bagging algorithm for the imputation of missing values in time series

Agung Andiojaya[a,b], Haydar Demirhan[b,*]

[a] Badan Pusat Statistik, Jl. Dr. Sutomo, 6 - 8, Jakarta 10710, Indonesia
[b] RMIT University, School of Science, Mathematical Sciences Discipline, 364 Swanston Street, Melbourne, Victoria 3000, Australia

ARTICLE INFO

Article history:
Received 4 December 2018
Revised 26 March 2019
Accepted 26 March 2019
Available online 27 March 2019

Keywords:
Block bootstrap
Gap filling
Interpolation
Kalman filter
Stineman
Weighted moving average

ABSTRACT

Classical time series analysis methods are not readily applicable to the series with missing observations. To deal with the missingness in time series, the common approach is to use imputation techniques to fill in the gaps and get a regularly spaced series. However, this approach has several drawbacks such as information and time bias, relationship causality, and not being suitable for the series with a high missingness rate. Instead of directly imputing the missing values, we propose a bagging algorithm to improve on the accuracy of imputation methods utilizing block bootstrap methods and marked point processes. We consider non-overlapping, moving, and circular block bootstrap methods along with amplitude modulated series and integer valued sequences. Imputation methods considered for bagging are Stineman and linear interpolations, Kalman filters, and weighted moving average. Imputation accuracy of the proposed algorithm is investigated by nearly 3000 yearly, quarterly, and monthly time series from different sectors under the "missing completely random" and "missing at random" missingness mechanisms. The results of the numerical study show that the proposed algorithm improved the accuracy of the considered imputation methods at most of the instances for different missingness rates and frequencies under both missingness mechanisms.

1. Introduction

Time series data is composed of consecutive observations taken at successive time points. There are two main types of time series data based on the length of intervals between the observation time points: regularly (evenly or equally) spaced time series and irregularly (unevenly or unequally) spaced time series. Analysis of regularly spaced series is straightforward by employing mainstream analysis methods. However, it is challenging for irregular series since the irregularity of the series causes loss of information and makes it harder to capture trend, serial correlation, and seasonality.

The gaps in the series occur when data can only be observed in certain times/periods (i.e. a patients health condition or a natural disaster), or a regular time series is considered as an irregular series when there are missing observations or unequal frequencies in multivariate series (Eckner, 2012). The mechanism behind missing observations is called Missing Completely at Random (MCAR), Missing at Random (MAR), or not Missing at Random (NMAR) (see Little & Rubin, 2014 for details). We need to know the reason of missingness for NMAR. Because we can impute missing values without knowing the reasons of missingness for both of MCAR and MAR, we focus on these two cases in this article.

The common approach to deal with the missingness caused by MCAR or MAR is to fill in the gaps by imputation methods, and then, apply the methods developed for regular time series for further analyses (Bermúdez, Corberán-Vallet, & Vercher, 2009; Gómez, Maravall, & Peña, 1999). Fung (2006a); Moritz, Sardá, Bartz-Beielstein, Zaefferer, and Stork (2015) and Lepot, Aubin, and Clemens (2017) provided comprehensive reviews of imputation methods for time series data. Moritz and Bartz-Beielstein (2015) published an R package called **imputeTS** for the implementation of main imputation methods to fill in the gaps in the univariate irregularly spaced series. Main imputation methods include linear/nonlinear interpolation, curve fitting, expectation-maximization, decompositions, bootstrapping, state space models, and Kalman filtering. Bokde, Beck, Álvarez, and Kulat (2018) proposed a novel imputation method based on capturing the repeating patterns of time series data. This approach utilizes pattern sequence forecasting algorithm that utilise periodic characteristics of time series data to produce forecasts. Bashir and Wei (2018) proposed a vector-auto-regressive imputation method composed of a hybrid of expectation minimization and prediction error

* Corresponding author.
E-mail address: haydar.demirhan@rmit.edu.au (H. Demirhan).

https://doi.org/10.1016/j.eswa.2019.03.044

Abbreviations

BB	Block bootstrap
NBB	Non-overlapping BB
MPP	marked point process
AMS	Amplitude modulated series
MSE	Mean squared error
DEM	Demographic
IND	Industrial
MIC	Micro economic
Impl.	Implementation
KF-AA	Kalman filter with auto-arima
KF-STS	Kalman filter with structural time series
WMA-Exp	Weighted moving average with exponential weights
MBB	Moving BB
CBB	Circular BB
BBM	Block bootstrap method
IVS	Integer valued sequence
MAE	Mean absolute error
FIN	Financial
MAC	Macro economic
OTH	Other
Stat	Error statistic
MCAR	Missing completely at random
MAR	Missing at random

minimization methods. Demirhan and Renwick (2018) recently compared performance of the methods covered in the **imputeTS** package for irregularly spaced solar radiation series and found that the linear and Stineman interpolations are very precise for minutely series, Kalman filters are very accurate for hourly series, and weighted moving average performs sufficiently for lower frequencies such as daily and weekly.

Instead of achieving the regularity of the series by imputation, another approach introduced by Jones (1962) utilizes spectral density estimation. As an extension of this approach, Parzen (1963) introduced amplitude modulated series (AMS) approach to analyze the series with missing observations. The AMS process builds an indicator to mark the gaps in the series. It has been applied to deal with the gaps in time series data under different conditions. Datta and Du (2012) investigated the application of AMS approach to estimate a heteroskedasticity and autocorrelation consistent covariance matrix on the irregularly spaced series. By comparing to the equally spaced model, they observed that the AMS approach performs better in the cases of small sample size and low autocorrelation, which can be seen for yearly series in practice.

To deal with irregularity, another marked point process technique called integer-valued sequence (IVS) was introduced recently by Niebuhr (2017). He proposed the construction of integer-valued sequences to mark the time points with observations and demonstrated the correspondence between the AMS and IVS approaches in a unique way. Niebuhr (2017) combined the AMS and IVS approaches with the moving block bootstrap (MBB) approach to account for the irregularity in the series, and proposed procedures to form the blocks by utilizing AMS and IVS modeling. He conducted a simulation study for the comparison of the accuracy of the MBB and AMS (MBB-AMS), MBB and IVS (MBB-IVS) combinations and dependent wild bootstrap (DWB) in the estimation of 5% and 90% percentiles of the finite-sample distribution of $\sqrt{n}\bar{X}_n$ over a simulated series X_n from AR(3) process. DWB is a block-free bootstrap technique introduced by Shao (2010).

Use of block bootstrap and AMS or IVS combinations does not ensure regularity of the series. The blocks formed by block bootstrap methods will still be irregular. Also, bootstrap is a well known method for evaluating the sampling variation of a statistic. In particular, block bootstrap is designed to preserve local dependencies in time series. Therefore, the direct use of block bootstrap to deal with irregularities in the series for modeling tasks that require regular/complete series is unsuitable.

The idea of bootstrap aggregation (bagging) proposed by Breiman (1996a) has been utilized for the imputation of missing values. Recently, Le and Benjapolakul (2018) considered bagging for imputation of missing synchrophasor data, which indicates the health of an electrical system. They use multiple linear regression model as the base model for bagging with observations recorded within very short time intervals. Considering the time dependency of observations even within a short time interval would improve the accuracy of the bagged regression estimators. Lee and Yang (2006, 2008) considered bagging binary predictors in time series data. They observe that bagging is useful in improving the accuracy of binary prediction in small sample but not in large samples where instability may decrease due to large sample size. Zheng et al. (2014) and Zheng (2018) used bagging predictors of posterior probabilities to impute genotypes of human leukocyte antigen. In these studies, bagging idea is used for multinomial cases further to bagging binary predictors. Application of block bootstrap would improve on the accuracy of the methods for genotype imputation since it samples neighbour single-nucleotide polymorphisms together. Kumutha and Palaniammal (2013) proposed a k-Nearest neighbor (kNN) approach with bagging for the imputation of microarray data. Working with microarray data in a stratified stetting helps reducing high computational cost due to the high dimensionality (Liu, Zhang, Gehan, & Clarke, 2002). In this sense, use of block bootstrap methodology over the strata created for a microarray dataset along with bagging would benefit in decreasing the computational cost for the imputation. Outside the imputation problem, bagging predictors has a great popularity in a wide range of real-world applications (see Beretta & Santaniello, 2016; Chatzis, Siakoulis, Petropoulos, Stavroulakis, & Vlachogiannakis, 2018; Foroozand & Weijs, 2017; Jain & Kanhangad, 2018; Liang, Zhu, & Zhang, 2011; Lin & Yin, 2016; Liu, Ouyang, & Li, 2017; Louzada, Anacleto-Junior, Candolo, & Mazucheli, 2011; Simidjievski, Todorovski, & Džeroski, 2015; Suzuki & Ohkura, 2016 for some of recent applications).

Bagging is based on building a classifier on each bootstrap sample generated from the learning sample and combining them to get an aggregated predictor. Bagging is very effective for the problems where a small change in the learning set can cause a notable change in the predictions (Breiman, 1996a; Breiman, 1996b). In our case, we do bootstrapping over a set of blocks created by using different block bootstrap methods and we have a time series imputation problem that involves predictors which are prone to the changes in local correlation structures in the series. We need to combine the predictors working under each block to get an aggregated, regular series for further modelling. In this paper, we propose an approach that employs bagging for the imputation of missing values in time series data along with a combination of block bootstrap and AMS or IVS marked point processes. We consider MBB, circular block bootstrap (CBB) introduced by Politis and Romano (1991), and non-overlapping block bootstrap (NBB) introduced by Carlstein (1986) along with the AMS and IVS approaches under the bagging framework. The utility of this approach is that while the block bootstrap approach is preserving the local dependencies within the series, AMS or IVS marked point process captures missingness patterns in the series. Because the missing value imputation methods are sensitive to the changes in the series, aggregating the imputations in each block by bagging improves imputation accuracy.

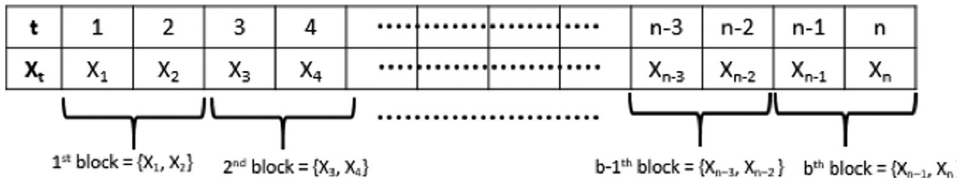

Fig. 1. Block configuration procedure for NBB technique with $\ell = 2$.

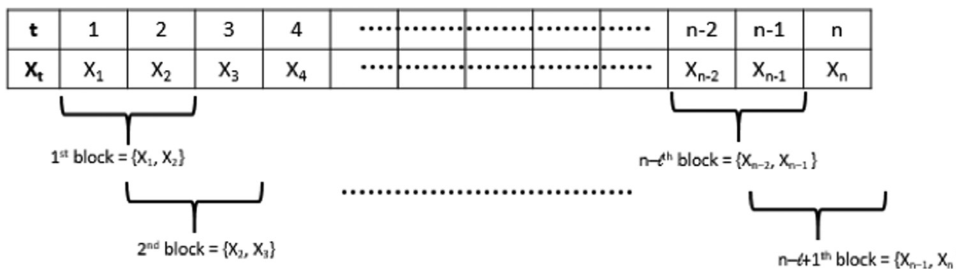

Fig. 2. Block configuration procedure for MBB technique with $\ell = 2$.

We conduct an extensive numerical study in a semi-Monte Carlo setting to assess the accuracy of the proposed bagging algorithm considering the MBB-AMS, MBB-IVS, CBB-AMS, CBB-IVS, NBB-AMS, and NBB-IVS combinations under different block lengths, missingness rates. We consider 2829 yearly, quarterly, and monthly industrial, financial, demographic, micro and macro economic, and some other real series in our numerical study. Missingness is randomly generated under low, moderate, and high missingness rates under the MCAR and MAR mechanisms.

2. Bagged block bootstrap implementation with marked point processes

2.1. Block bootstrap methods

Suppose we have observed the series $\boldsymbol{X} = \{X_t : t = 1, 2, \ldots, n\}$ which includes missing observations. The block bootstrap method preserves the structural dependency within a pseudo-sample time series. The block bootstrap method has been developed independently by Carlstein (1986); Hall (1985) and Kunsch (1989) from the initial idea of developing criteria to create blocks of consecutive data. The procedure begins with splitting the original series into blocks with block length (the number of consecutive observations in each partition) equal to ℓ. Then, b blocks are generated, n/ℓ blocks are drawn with replacement from the set of the generated blocks, and the resampled blocks are arranged into a new bootstrap series. There are two ways of forming the blocks: non-overlapping and overlapping formation.

2.1.1. Non-overlapping block bootstrap

Carlstein (1986) developed the NBB to form a block bootstrap approach with non-overlapping partitions of data embodied in the sequence of the blocks. In NBB, the block length is specified for a series of length n based on the following:

$$\ell \equiv \ell_n \in [1, n], \tag{1}$$

where ℓ is an integer. Then, the number of blocks b can be obtained by the relation $\ell b \le n$ where $b \ge 1$ and b is the largest integer that fulfill Eq. (1). This procedure produces the blocks $B_i = (X_{(i-1)\ell+1}, \ldots, X_{i\ell})$ for $i = 1, \ldots, b$. An illustration of the NBB procedure to configure the blocks with $\ell = 2$ is given in Fig. 1.

After the procedure to form the blocks is completed, we draw k blocks randomly with replacement from the set $\{B_1, B_2, \ldots, B_b\}$ such that k is the smallest integer satisfies

$$\ell k \ge n. \tag{2}$$

2.1.2. Overlapping block bootstrap

Kunsch (1989) and Liu and Singh (1992) independently introduced the overlapping block bootstrap which is also known as MBB. The number of blocks generated from MBB is $b = n - \ell + 1$. Fig. 2 illustrates the MBB procedure.

After the set $\{B_1, B_2, \ldots, B_{n-\ell+1}\}$ is generated, we randomly select k blocks with replacement such that k is the smallest integer satisfies Eq. (2) (Politis & Romano, 1991).

2.1.3. Circular block bootstrap

Although MBB is a widely used method, there is a bias caused by the block configuration procedure that leads to the reduced weight of X_i's when $i < \ell$ or $i > n - \ell + 1$. For instance, when $\ell = 2$, X_1 and X_n are only included once in a block. Whereas the other observations are included in 2 blocks (see Fig. 2). Thus, according to Politis and Romano (1991), there are different weights of X_i causing the series need to be 'wrapped' in a circle in order to get an unbiased bootstrap distribution. The overlapping block bootstrap method addressing this issue is known as CBB. Basically, the block formation of CBB is similar to that of MBB. But, in the CBB, the last block covers another bunch of $\ell - 1$ of the first observations. This procedure is summarized in Fig. 3.

After the set $\{B_1, B_2, \ldots, B_n\}$ is generated, we randomly select k blocks with replacement such that k is the smallest integer satisfies Eq. (2) (Politis & Romano, 1991).

For all the block bootstrap methods discussed in this section, selection of the optimal block length is another critical consideration. We discuss this issue along with the design of our simulation space in Section 3.

2.2. AMS and IVS approaches

For the AMS form of the original series, we define a missing value indicator a_t such that

$$a_t = \begin{cases} 1, & \text{if the value of } X_t \text{ is observed at } t, \\ 0, & \text{if the value of } X_t \text{ is missing at } t. \end{cases} \tag{3}$$

Thus, the sequence a_t marks the missing observations and the number of non-zero observations can be stated as $m = \sum_{t=1}^{n} a_t$.

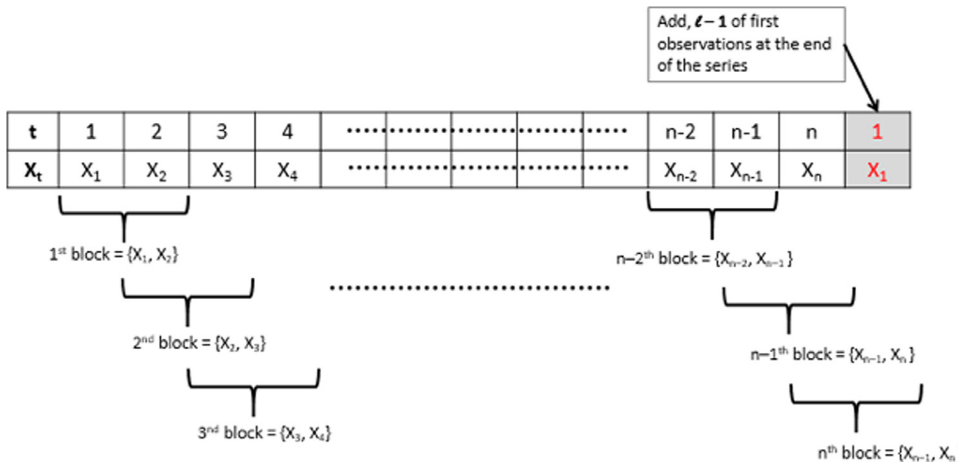

Fig. 3. Block configuration procedure for CBB technique with $\ell = 2$.

In the IVS approach, observations are marked as

$$(X_{\sum_{i=1}^{1} Z_i}, X_{\sum_{i=1}^{2} Z_i}, \dots, X_{\sum_{i=1}^{m} Z_i}), \tag{4}$$

where $\boldsymbol{Z} = \{Z_t : t = 1, 2, \dots, m\}$ is a strictly positive and integer-valued time series that define the index of observed elements of an irregular series. Suppose we have the series $\{X_1, X_2, X_3^*, X_4^*, X_5, X_6^*, X_7\}$ where $X_\cdot^*$ shows a missing observation. Then, we respectively obtain the following by Eqs. (3) and (4):

$$\{a_1, a_2, a_3, a_4, a_5, a_6, a_7\} = \{1, 1, 0, 0, 1, 0, 1\}$$

and

$$\{Z_1, Z_2, Z_3, Z_4, Z_5, Z_6, Z_7\} = \{1, 1, 0, 0, 3, 0, 2\}.$$

Practically, in the AMS approach, we replace the missing observations by 0 and the length of the new AMS series remains n. Whereas, in the IVS approach, we just skip the missing observations using the observation process $\boldsymbol{Z}$; and hence, the length of new IVS series becomes m. Gómez et al. (1999) use a similar skipping approach to fit ARIMA models before interpolating for the missing observations.

Niebuhr (2017) outlines the details of forming the blocks under MBB method in combination with AMS and IVS approaches. He also demonstrates bootstrap validity of the MBB-IVS combination. In the block bootstrap implementation with the AMS and IVS approaches, the length of computed bootstrap series is n and m, respectively. When the sampler hits to a missing observation, the corresponding bootstrap replication is 0 under the AMS approach and missing observations are skipped while forming the bootstrap replications under the IVS approach.

Instead of using the bootstrap samples coming out of the AMS and IVS approaches, we propose to utilise the AMS and IVS approaches to build the blocks for the mentioned block bootstrap methods in Section 2.2. In our approach, we neither replace missing observations with zero nor skip them in the analysis. Instead, we utilize the AMS and IVS sequences to specify the starting and end points of the blocks and keep the missing observations as is. By this approach, we include the information on the missingness pattern in the bootstrap method by using the AMS and IVS.

2.3. Bagging block bootstrap imputation

Suppose the learning set $\mathcal{L}$ is composed of the stationary series $\{X_t, t = 1, \dots, n\}$ with missing observations and $\varphi(X_t, \mathcal{L})$ is the predictor for the missing values in the series. In the block bootstrap approach, we sample with replacement over the blocks $\{B_1, B_2, \dots, B_b\}$, where b is specified by the block formation method, to create each bootstrap replication. With N replications, we have a sequence of N learning sets $\{\mathcal{L}^{(B)}\}$ composed of the elements of repeated bootstrap samples $\{\{B_{11}^*, B_{21}^*, \dots, B_{b1}^*\}, \dots, \{B_{1N}^*, B_{2N}^*, \dots, B_{bN}^*\}\}$. Then, we apply an imputation method with each of $\{\mathcal{L}^{(B)}\}$ to get the predictor $\varphi(X_t, \mathcal{L}^{(B)})$ of missing values in the series. Then, following Breiman (1996a), we obtain aggregated predictor $\varphi_B(X_t)$ as $av_B\varphi(X_t, \mathcal{L}^{(B)})$.

Implementation of the proposed approach is given by Algorithm 1.

Algorithm 1 Bagged block bootstrap imputation

1. Given the irregular series $\boldsymbol{X} = \{X_t : t = 1, 2, \dots, n\}$, configure and generate the blocks for bootstrapping by employing the AMS approach over Eq. (3) as shown in Fig. 4 or IVS approach over Eq. (4) as shown in Fig. 5 in combination with one of NBB, MBB, or CBB methods.
2. Set $i = 1$, randomly select b blocks with replacement from the set of generated blocks at step 1, and arrange it into a new series $\boldsymbol{B}_i^* = \{B_1^*, B_2^*, \dots, B_b^*\}$, which still includes missing observations.
3. Apply an imputation method to $\boldsymbol{B}_i^*$ to get a regular series $\boldsymbol{X}_i$.
4. Set $i = i + 1$ and repeat the steps 2 and 3 until $i \geq N$.
5. Compute row-wise average of the matrix $\boldsymbol{S}_{n\times N} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_N)$ to do the aggregation and obtain the regular series for further analyses.

At the first step of Algorithm 1, the procedure to form the blocks with the AMS technique is similar to the ordinary block bootstrap where each block is formed by the partition of ℓ consecutive values in the series. In total, n/l, $n - \ell + 1$, and n blocks are produced by the NBB, MBB, and CBB methods, respectively. The key point in the block bootstrap part of our approach is that the block bootstrap generates blocks with fixed length of ℓ. The illustration of formation of the blocks at step 1 of Algorithm 1 by using the AMS approach for non-overlapping and overlapping block bootstrap methods is shown in Fig. 4.

In the IVS approach, partitioning the series into blocks is based on the observed values. Hence, the block length ℓ refers to the number of observed values. The implication of this implementation is that every block has fixed ℓ number of observed values, but the length between blocks varies. The formation of blocks with the IVS approach at step 1 of Algorithm 1 for non-overlapping and overlapping block bootstrap is illustrated in Fig. 5.

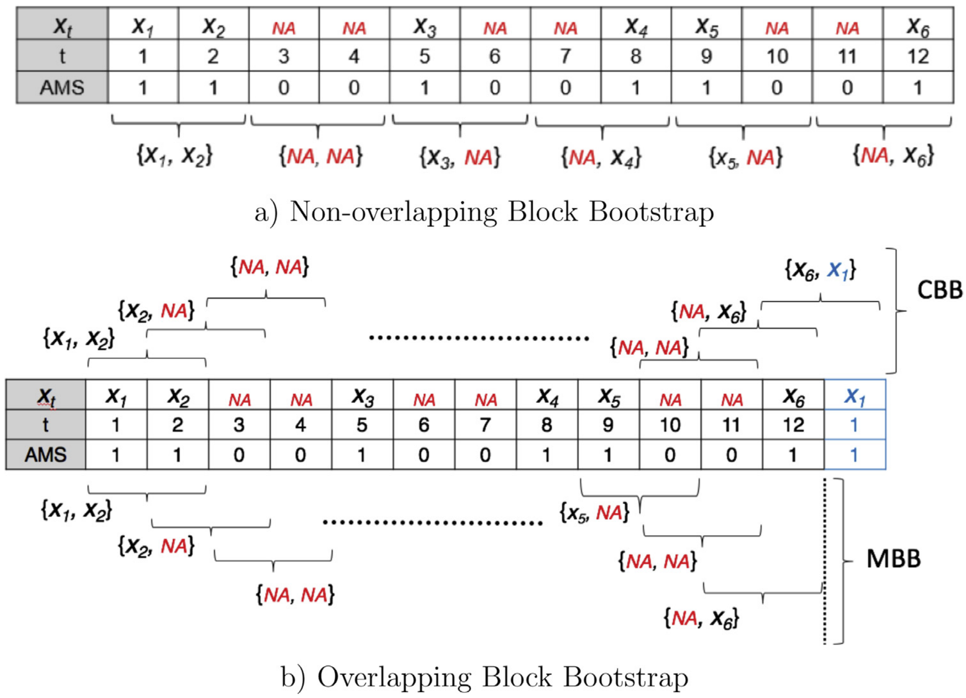

Fig. 4. Block configuration for non-overlapping block bootstrap (panel a) and overlapping block bootstrap (panel b) using the AMS approach with $\ell = 2$.

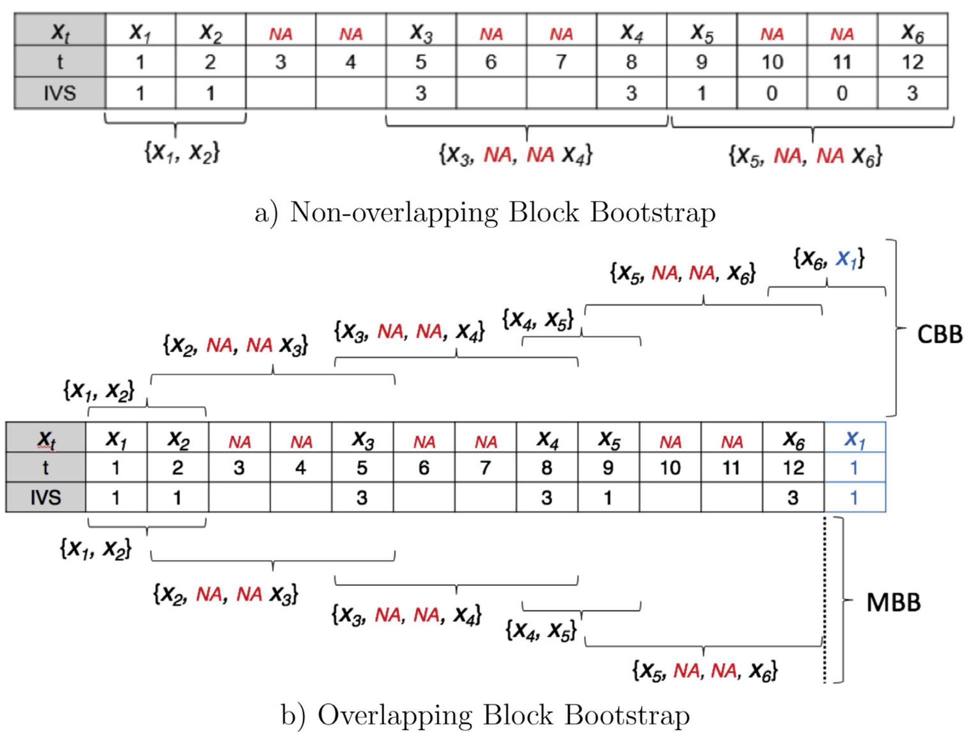

Fig. 5. Block configuration for non-overlapping block bootstrap (panel a) and overlapping block bootstrap (panel b) using the IVS approach with $\ell = 2$.

The AMS and IVS utilization affects the block configuration of block bootstrap methods. The combination of block bootstrap and AMS yields blocks that contain a fixed number of series. Whereas the block bootstrap and IVS combination produces blocks that contain a fixed number of observed values. According to these block generating rules, block bootstrap based on AMS produces b blocks that have fixed block length, and block bootstrap based on IVS produces b blocks that have varying lengths. By replicating the bootstrap procedure N times, N bootstrap samples are generated, missing values in each replication is imputed by an imputation method, and final aggregation is done by taking the row-wise average of the regularized bootstrap replications at step 5 of Algorithm 1. Algorithm 1 is summarized in Fig. 6.

3. Numerical study

We conduct a comprehensive semi-Monte Carlo (semi-MC) simulation study with real data to assess the imputation accuracy of the proposed bagging algorithm and benchmark its accuracy with the mainstream imputation methods. Details of the semi-MC

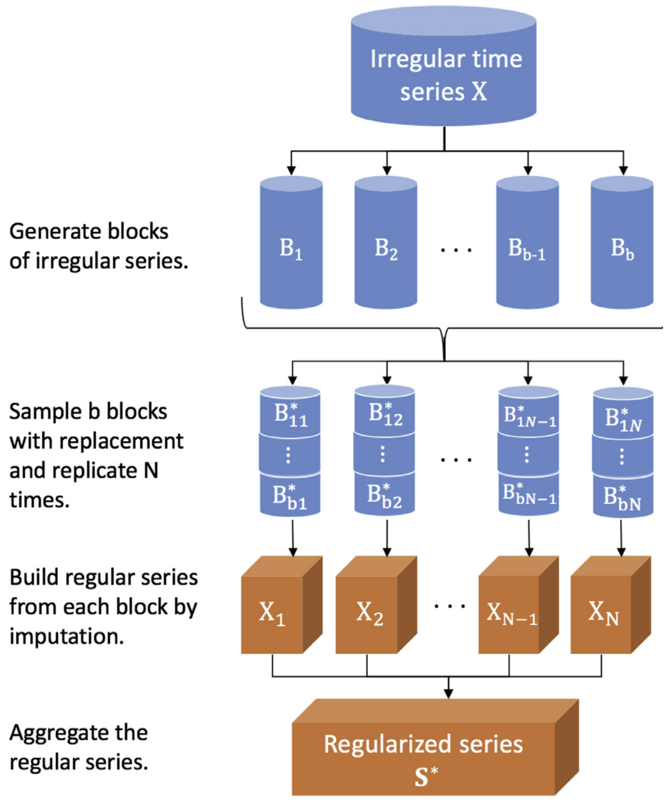

Fig. 6. Bagged block bootstrap imputation of missing values in time series.

simulation study and its results are articulated in the following subsections.

3.1. Simulation space and implementation

We benchmark the accuracy of our approach with the regular implementation of five imputation methods: Stineman interpolation (Stineman, 1980), Kalman filters with smoothing and auto-arima (KS-AA), Kalman filters with smoothing and structured time series (KS-STS), linear interpolation, and exponentially weighted moving average (WMA-Exp) (see Fung, 2006b and Demirhan & Renwick, 2018 for the details of these methods). Demirhan and Renwick (2018) use window lengths of 2, 4, and 6 with WMA-Exp method and do not observe a considerable difference in the accuracy of imputations for lower frequencies such as weekly. By referring this result, we set the window length to 4 in our simulations. For Kalman filters, a structural model is fitted by maximum likelihood method and the state space representation of ARIMA model (Grewal & Andrews, 2008). These imputation methods are widely used for the imputation of missing values and supported by the comparative studies in the literature (Demirhan & Renwick, 2018). To mimic real-life situations, we used a dataset that includes 2829 yearly, quarterly, and monthly real time series. This dataset was used for a forecasting competition called M3-Competition (Makridakis & Hibon, 2000). The M3-Competition dataset (available at https://forecasters.org/resources/time-series-data/m3-competition) provides a sufficient coverage of real situations with its contents including micro and macro economic series, finance series, demographic series, those from industry, and some other series out of these classifications. Table 1 shows the classification of M3 series by their frequency and category of the series (Makridakis and Hibon, 2000, p. 454).

The histograms of sample sizes for each type and frequency of the series are given in Fig. 7. While industry series have the maximum lengths, micro economics series have the shortest lengths in the M3 dataset.

To further describe the dataset, we display box-plots of autocorrelations at the first 20 lags for ordinary differenced yearly series, 20 lags for ordinary and seasonally first differenced quarterly series, and 40 lags for ordinary and seasonally first differenced monthly series in Figs. A1, A2, and A3 of Appendix A, respectively. In these figures, each box-plot shows the distribution of autocorrelations among each series type for each given lag. For monthly series, we display autocorrelations up to 40 lags in Fig. A3 due to having longer series for the monthly frequency (Fig. 7). For quarterly and monthly series, since we observe seasonal trend for all series types, the first seasonal difference is applied in addition to the first ordinary difference.

For yearly series, each series type has a different autocorrelation characteristic. Macro economic series have mostly positive first lag autocorrelation. Micro economic series has both negative and positive first lag autocorrelation in wide rage of possible values. The median first lag autocorrelation is close to zero for industrial yearly series. Finance series show large autocorrelations at higher lags as well as the first lag. Yearly demographic series exhibit wider ranges of autocorrelations at the first several lags suggesting dependence to two-three previous years and changing serial dependence structure after several time points. This would be creating sudden slope changes in the series.

For quarterly frequency, macro economic series has relatively lower first lag autocorrelations within a wide range of values. Their first seasonal lag autocorrelation is also slightly lower than the other series types. Finance series have high first lag autocorrelations at the first ordinary and seasonal lags and the autocorrelations at the second seasonal lags are mostly close to zero. Quarterly demographic series have high autocorrelations at first several ordinary lags and their first seasonal lag autocorrelation is lower than that of the other series types.

For industrial, finance, and macro economic monthly series, variation of autocorrelations at seasonal lags are higher than micro economic series. Micro economic and industrial series have mostly high negative first lag autocorrelations. Finance series have both positive and negative first lag autocorrelations. Demographic series have high autocorrelations at first several lags suggesting longer terms of dependence than the other series types. For each frequency, each series type provides a different ordinary and seasonal autocorrelation characteristic; and hence, they provide a wide variety of experimental series for our numerical study.

In our semi-MC approach, we generate missing values within each of the M3 series based on the MCAR and MAR mechanisms, and then, compare our algorithm and benchmark methods in terms of imputation accuracy. Missing values are generated by using the R package **imputeTestbench** (Beck & Bokde, 2017) for the MAR case and completely randomly for MCAR case. Fig. 8 shows the simulation space and summarizes the semi-MC implementation.

First, we ensure that each series is stationary as required by the block bootstrap methods by using the augmented Dickey-Fuller unit root test and apply differencing if necessary. In practice, straightforward tests of stationarity are not applicable in the presence of missing values. For this case, Ryan and Giles (1998) outline testing for a unit root when the series is irregular or having missing observations. After testing for stationarity, we generate missing values in each series by using a random binomial process with the missingness rates of $p = 0.1, 0.25,$ and 0.5, which respectively create low, moderate, and high rates of missingness in the series.

The optimal block length for the considered block bootstrap methods is decided based on the previous studies in the literature. Carlstein (1986) considered minimizing the mean square error (MSE) to determine the optimal block length. Kunsch (1989)

Table 1
The number of M3 series under each series type and frequency.

Frequency[a]	Micro	Industry	Macro	Finance	Demographic	Other	Total
Yearly	146	102	83	58	245	11	645
Quarterly	204	83	336	76	57	0	756
Monthly	474	334	312	145	111	52	1428
Total	824	519	731	279	413	63	2829

[a] The original competition dataset includes another bunch of 147 series with other frequencies than those mentioned in the table. These series are not included in our study due to their undefined frequency.

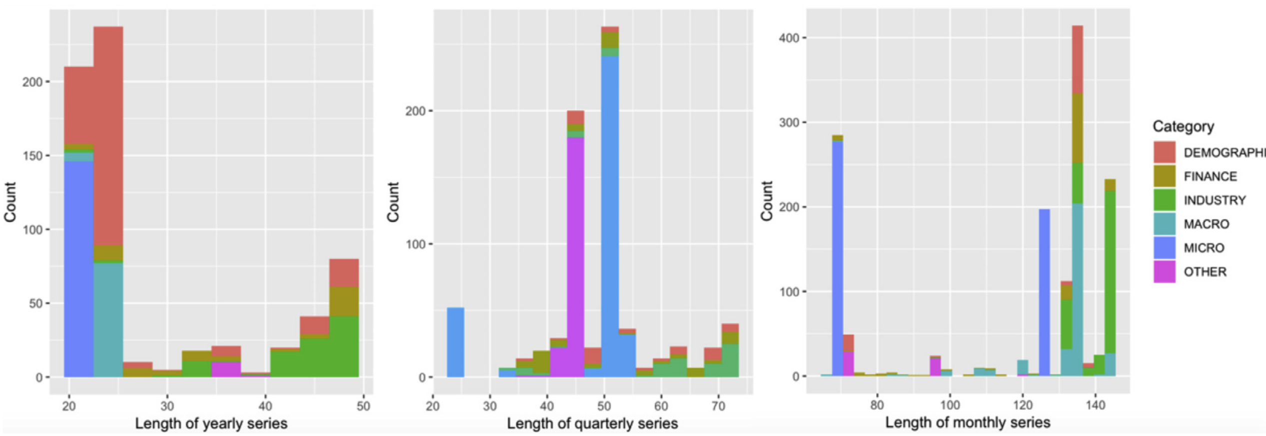

Fig. 7. Histograms of sample sizes for frequency and type of M3 series.

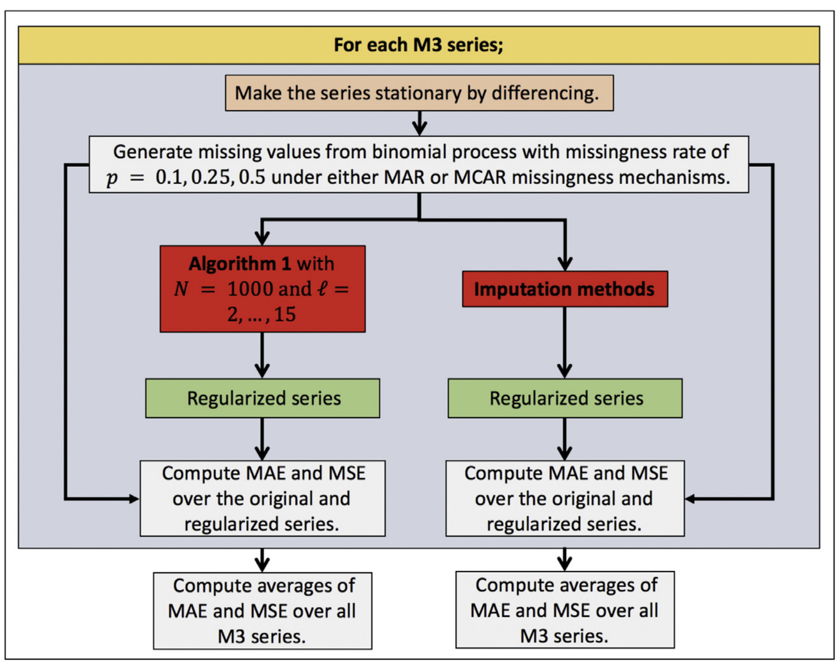

Fig. 8. The semi-MC simulation space and its implementation to compare the accuracy of the considered imputation methods with and without proposed bagging approach.

suggested that we get the optimal block length if the number of blocks is proportional to $n^{1/3}$. The implication of Kunsch (1989)'s rule is having longer block lengths. Another study conducted by Hall and Horowitz (1993) evaluated the existing block length formulations. They found that the optimal block length rule for moving block bootstrap scheme (they named it as Kunsch's rule) and the non-overlapping block scheme (they named it as Carlstein's rule) are similar. Further, they reformulated the rule as:

$$\ell^* = \begin{cases} (1.5n)^{1/3}|\rho|^{-2/3}, & \text{under Kunsch's rule,} \\ n^{1/3}|\rho|^{-2/3}, & \text{under Carlstein's rule,} \end{cases} \tag{5}$$

where ρ is calculated as in Eq. (6) for ARMA(1,1) process:

$$\rho = \frac{(\theta+\phi)(1+\theta\phi)}{1+2\theta\phi+\phi^2}. \tag{6}$$

Using Eqs. (5) and (6) for the considered combinations of the ϕ, θ and p without $(\phi, \theta) = (-0.9, 0.9), (0.9, -0.9), (-0.1, 0.1), (0.1, -0.1)$, we get the maximum of the optimal block length $\ell^*_{max} \approx 19$ when $n = 200$ and $p = 0, 0.1$, and $\ell^*_{max} \approx 18$ when $n = 200$ and $p = 0, 0.25$ under Kunsch's rule while the minimum is $\ell^*_{min} \approx 4$ when $n = 100$ and $p = 0.5$ under Carlstein's rule. But, we have $\ell^*_{max} \approx 15$ over the rest of the combinations of n, p, ϕ, and θ. Because we cover a wide range of possible series in practice with these settings of ARMA(1,1) process, we set the block length as $\ell^* = 2, 5, \ldots, 15$ for all combinations of MBB, NBB, and CBB block bootstrap methods and AMS/IVS approaches in our numerical study. In total, we apply 6 block bootstrap - marked point process combinations at the second stage shown in Fig. 6.

From 1000 bootstrap samples generated through bootstrapping stage, the new regularly spaced series is created by Algorithm 1 with the specified block lengths. Then, the imputed values of artificially created missing values are compared to their true values over M3 series using mean absolute error (MAE; Hyndman, Koehler, Ord, & Snyder, 2008, see p. 25 and MSE Hyndman et al., 2008, see p. 25) to assess the gain in accuracy by the proposed bagged imputation algorithm. For the regular implementation of the considered imputation methods, we used the R package **imputeTS** version 2.6 (Moritz & Bartz-Beielstein, 2015).

3.2. Simulation results

3.2.1. Yearly series

For yearly series, average MSE values of bagged and regular versions of imputation methods are given in the breakdown of the category of series and missingness rates for the MCAR case by Fig. 9 and for the MAR case by Fig. C1 of Appendix C. Each point in each panel of Fig. 9 and Figure C1 corresponds to a combination of block bootstrap method and marked point process; hence, there are 6 points for each color and shape combination. The 45-degrees line is imposed on each panel of the figures to help the comparison of MSE values. In the rest of the MAE and MSE plots in the article, we use the same setting unless otherwise mentioned.

For all of the considered imputation methods except some cases with the Kalman filters with auto-arima, the proposed bagging algorithm gives more accurate imputations in terms of MSE for nearly all combinations of block bootstrap methods (BBMs) and marked point processes (MPPs) for both of the MCAR and MAR cases. Both bagged and regular implementations of imputation methods are not dramatically affected by the missingness rate. As expected, MSE values are slightly smaller with 10% missing rate for all methods.

For the Kalman filters with auto-arima, the bagging algorithm has higher MSE values than the regular implementation for some combinations of BBM and MPP over the series in the “other” category. However, we get better MSE values for bagged implementation of the Kalman filters with auto-arima under non-overlapping block bootstrap with AMS approach as can be seen in Table D1 in Appendix D.

MAE values for each imputation method, series type, and missingness rate are shown in Fig. B1 of Appendix B for the MCAR case and in Fig. C2 of Appendix C for the MAR case. We have similar MAE results to MSEs for yearly series for both of the MCAR and MAR cases. For the series in the “other” category, some MAE values for some BBM-MPP combinations are above the 45-degrees line. However, bagged implementation with NBB gives lower MAE values for all imputation methods (see Tables D1 and D4). Since blocks are not overlapping in NBB, it keeps the information on the time dependency of observations in the series in a more accurate way during imputations. Overall, we can conclude that the proposed algorithm significantly improves the imputation accuracy of the considered imputation methods for yearly series.

For the comparison of imputation methods, Table 2 presents minimum MSE and MAE values averaged over missingness rates and BBM-MPP combinations for bagged implementations, and average MSE and MAE values for regular imputation methods over the missingness rates under the MCAR case. The same information for the MAR case is presented by Table C1 of Appendix C. We take minimum over BBM and MPP for bagged algorithm because the accuracy of methods differs for BBM-MPP combinations, and the straightforward practice is to go for the best BBM-MPP combination. The proposed bagging implementation with Stineman interpolation and Kalman filters with auto-arima gives the lowest MAE and MSE values for all types of series for both of the MCAR and MAR cases. Stineman interpolation performs better than Kalman filters for yearly demographic series (in terms of both MAE and MSE) and micro economic series (in terms of MAE). Stineman interpolation is meant to perform sufficiently with the cases having sudden slope changes and volatility clustering. By our descriptive analysis of MC3 dataset, we expect sudden slope changes for demographic series; and hence, we consistently get better results with Stineman interpolation for the demographic series. For micro economic series, it is straightforwardly expected to have volatility clustering for which Stineman interpolation performs better. When Kalman filters applied with an ARIMA transition model, they reflect the autocorrelation structure of the observed part on the imputations over an ARIMA model identified by minimizing the corrected Akaike Information Criterion. Capturing the autocorrelation structure of the observed part makes Kalman filters with auto-arima more accurate than the other compared imputation methods.

For yearly series, we observe smaller MAE and MSE values by the proposed bagging implementation under the MAR case than those under the MCAR case. Under the MAR case, gaps in time series occur as continuous blocks under the autocorrelation structure inherent with the series. The block bootstrap implementation in our approach captures the missingness patterns along with the associated autocorrelation structures when longer gaps are created by the MAR mechanism rather than shorter gaps or individual missing values created by the MCAR mechanism.

To compare the accuracy of BBM-MPP combinations, Tables 3 and C2 show MAE and MSEs of bagged imputation methods for each combination of BBM and MPP under the MCAR and MAR cases, respectively. For yearly series, the lowest MAE and MSE values are observed for the bagging with non-overlapping block bootstrap with AMS implementation. Also, non-overlapping block bootstrap implementation with IVS gives the second best error values under all imputation methods. Circular block bootstrap and moving block bootstrap implementations have close MAE and MSE values to each other for all imputation methods. While AMS flags the missing values, IVS skips them. When combined with overlapping BBMs such as CBB and MBB, IVS does not keep original information on the time order of the missing observations. Therefore, the imputation with IVS and overlapping BBMs would require longer block lengths to give as accurate results as those combined with AMS.

To asses the effect of block length on the accuracy of the proposed bagged imputation algorithm for yearly series, histograms of block lengths giving minimum MAE value for each series type for Stineman interpolation and Kalman filtering with auto-arima methods are shown in Fig. 10 and Fig. C3 for the MCAR and MAR cases, respectively. Since the results for the other imputation methods are very similar to Fig. 10, they are omitted here. Considering that the length of yearly series in M3 dataset is mostly short (around 20), we get minimum error values around a short block length of 2 for all series types under both of the MCAR and MAR cases. Considering that yearly series mostly have shorter terms of serial dependency (a high first lag autocorrelation diminishing very quickly), getting more accurate results for short block lengths is a consistent result with the structure of the yearly series.

3.2.2. Quarterly series

The scatter plot of average MSE values for bagged and regular imputations are given by Fig. 11 and Fig. C4 of Appendix C for the MCAR and MAR cases, respectively along with the 45-degrees

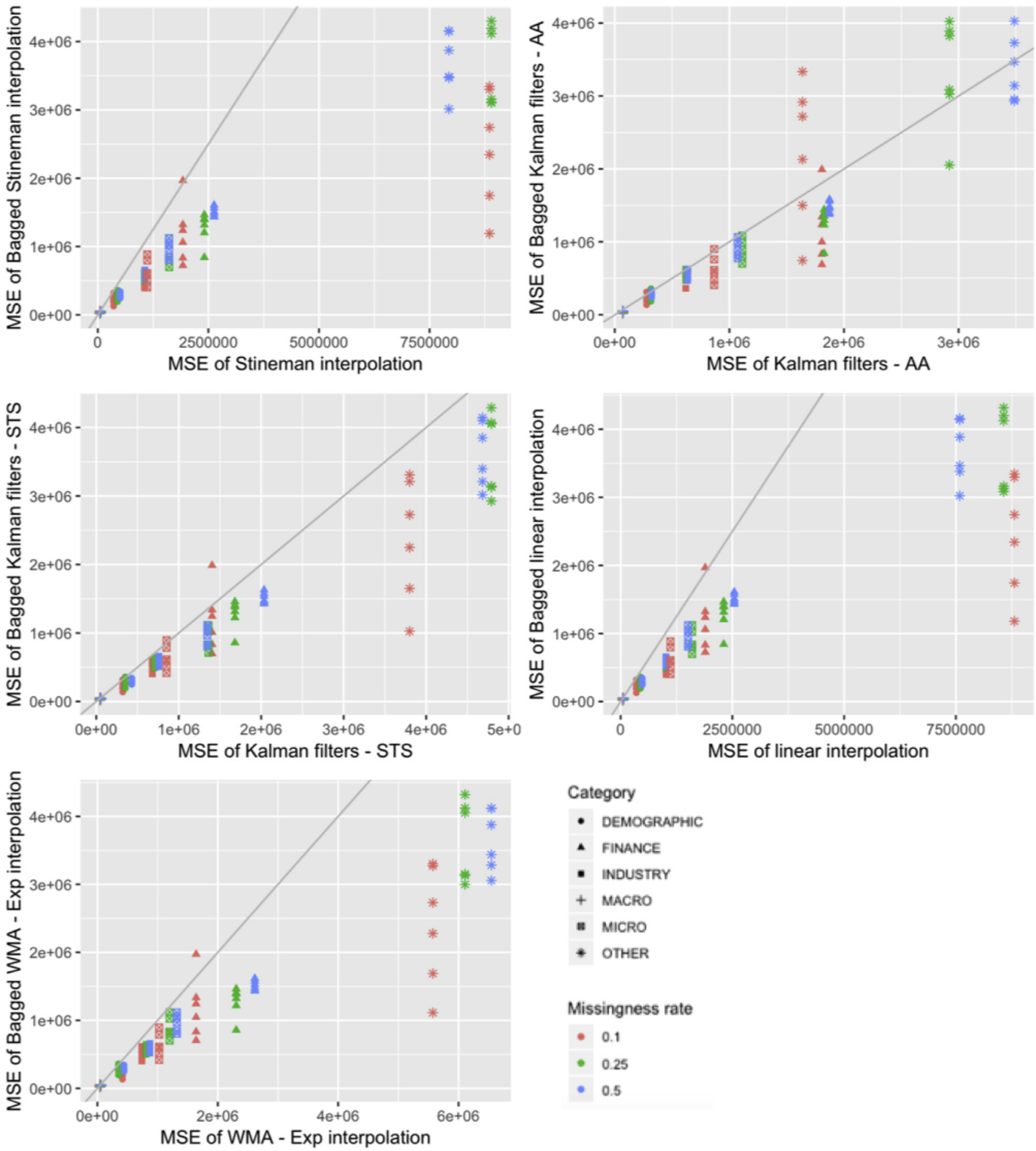

Fig. 9. Comparison of MSE values bagged and regular versions of interpolation methods for yearly series. The MSE values are averaged over BBM-MPP combinations for the bagged implementation. The lines on the panels show the 45-degrees regression line for comparison.

line for comparison. For both of the missingness mechanisms, the proposed bagging algorithm improves on the accuracy of the imputation methods for quarterly series as well. For micro economic, demographic, and industrial series, the bagged version of Kalman filters with auto-arima does not give lowest MSE values for all of the BBM-MPP combinations. However, we get the lowest MSE values for bagged implementation of Kalman filters with auto-arima under the NBB-AMS combination (see Tables D2 and D5).

The MAE values of the bagged and regular implementations of imputation methods are plotted in Fig. B2 of Appendix B and Fig. C5 of Appendix C for the MCAR and MAR cases, respectively. MAE results are very similar to MSE results in Fig. 11 for the MCAR case. In terms of MAE, we get better results under the MAR case with the proposed bagging algorithm. Bagged implementations give less MAE values for all methods, series types, and missingness rates (refer to Tables D2 and D5 for the detailed results in the breakdown of the BBM-MPP combinations).

For the comparison of imputation methods in quarterly series, average minimum MSE and MAE values over missingness rates and the BBM-MPP combinations for bagged implementations and

Table 2
Average minimum MSE and MAE values for bagged implementation and average MAE and MSE values for regular implementation of imputation methods over block bootstrap methods and marked point processes for yearly series under the MCAR case. The lowest MAE and MSE values for each series type is highlighted in boldface.

		Stineman		KF-AA		KF-STS	
Stat.	Category	Bagged	Regular	Bagged	Regular	Bagged	Regular
Min. MAE	DEM	**159.3**	287.3	169.9	282.3	168.8	274.8
	FIN	477.4	705.7	**475.1**	703.9	487.9	656.8
	IND	351.3	586.0	**338.2**	478.2	347.3	487.6
	MAC	74.0	132.8	**71.7**	177.1	**71.7**	132.7
	MIC	**322.7**	570.2	335.7	548.7	331.5	516.4
	OTH	811.1	2070.1	**588.4**	919.1	753.3	1444.4
Min. MSE	DEM	**1.29E+05**	3.56E+05	1.38E+05	2.77E+05	1.39E+05	3.21E+05
	FIN	7.24E+05	1.92E+06	**6.89E+05**	1.81E+06	6.99E+05	1.40E+06
	IND	3.96E+05	1.05E+06	**3.63E+05**	6.19E+05	4.03E+05	6.80E+05
	MAC	1.83E+04	4.72E+04	**1.57E+04**	6.93E+04	1.71E+04	4.43E+04
	MIC	4.04E+05	1.12E+06	4.06E+05	8.67E+05	4.16E+05	8.52E+05
	OTH	1.19E+06	7.94E+06	**7.45E+05**	1.64E+06	1.03E+06	3.80E+06?

		Linear		WMA-Exp	
Stat.	Category	Bagged	Regular	Bagged	Regular
Min. MAE	DEM	160.1	284.3	167.3	275.7
	FIN	478.8	700.4	494.3	682.7
	IND	350.9	573.6	351.7	503.2
	MAC	73.8	132.9	74.1	133.1
	MIC	322.8	565.9	331.1	520.5
	OTH	809.1	2027.4	792.5	1711.5
Min. MSE	DEM	**1.29E+05**	3.51E+05	1.35E+05	3.50E+05
	FIN	7.25E+05	1.89E+06	7.04E+05	1.64E+06
	IND	3.96E+05	1.01E+06	4.02E+05	7.39E+05
	MAC	1.83E+04	4.70E+04	1.85E+04	4.13E+04
	MIC	**4.03E+05**	1.11E+06	4.17E+05	1.03E+06
	OTH	1.18E+06	7.59E+06	1.11E+06	5.57E+06

Table 3
Overall averages of MSE and MAE values of bagged imputation for each combination of block bootstrap methods and marked point processes for yearly series under the MCAR case.

Stat	BBM	MPP	Stineman	KF-AA	KF-STS	Linear	WMA-Exp
MAE	CBB	AMS	551.7	536.6	551	552.2	551.7
	MBB	AMS	546.7	537.0	547.2	547.1	547.9
	NBB	AMS	**475.3**	**447.6**	**469.7**	**474.9**	**475.3**
	CBB	IVS	565.7	551.7	563.9	566.0	565.2
	MBB	IVS	565.6	555.7	564.6	565.7	564.7
	NBB	IVS	539.2	532.2	538.4	539.2	539.2
MSE	CBB	AMS	1.06E+06	9.90E+05	1.05E+06	1.06E+06	1.05E+06
	MBB	AMS	1.11E+06	1.06E+06	1.10E+06	1.11E+06	1.11E+06
	NBB	AMS	**8.85E+05**	**7.73E+05**	**8.50E+05**	**8.78E+05**	**8.67E+05**
	CBB	IVS	1.11E+06	1.06E+06	1.10E+06	1.11E+06	1.11E+06
	MBB	IVS	1.10E+06	1.07E+06	1.10E+06	1.10E+06	1.10E+06
	NBB	IVS	1.02E+06	9.85E+05	1.02E+06	1.02E+06	1.02E+06

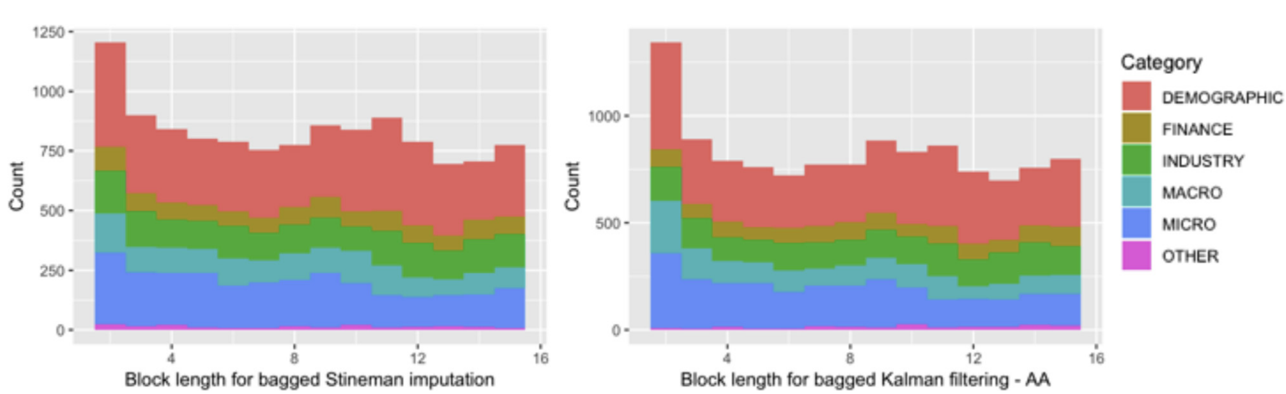

Fig. 10. Histogram of block lengths giving minimum MAE value for each series type for Stineman interpolation and Kalman filtering with auto-arima methods over all yearly series.

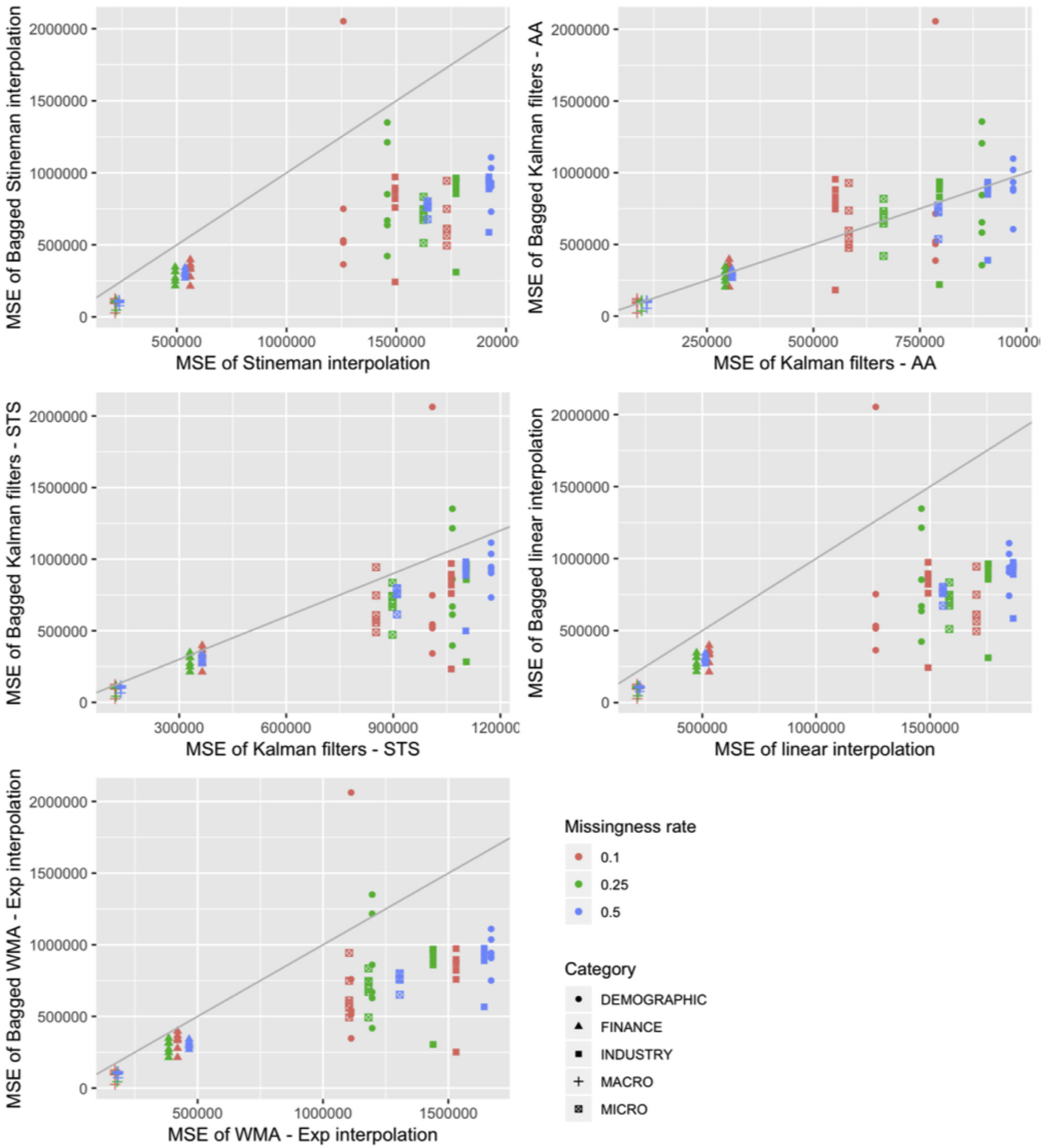

Fig. 11. Comparison of MSE values bagged and regular versions of interpolation methods for quarterly series. The MSE values are averaged over BBM-MPP combinations for the bagged implementation. The lines on the panels show the 45-degrees regression line for comparison.

average MSE and MAE values over the missingness rates for regular imputation methods are given in Table 4 and Table C3 of Appendix C for the MCAR and MAR cases, respectively. In terms of the best performing imputation method with bagging, the results are very similar to each other under the MCAR and MAR cases. For quarterly series, Stineman interpolation is not as accurate as in the yearly series. A possible reason of this observation is having changing slopes and seasonality in a confounded way in quarterly series. Kalman filters with auto-arima has the lowest MSE and MAE values with bagged implementation for least one of the BBM-MPP combinations. Similar to the case with yearly series, the auto-arima option captures the seasonal serial dependency in addition to the ordinary autocorrelation structure and we have a wide variety of serial correlation structures among the series types in our dataset. As the result of capturing the serial dependence structure appropriately, we get more accurate results with Kalman filters with auto-arima method.

To figure out the BMM-MPP combination giving the lowest error values, we tabulated MAE and MSE values in the breakdown of the BBM-MPP combinations in Table 5 and Table C4 of Appendix

Table 4
Average minimum MSE and MAE values for bagged implementation and average MAE and MSE values for regular implementation of imputation methods over block bootstrap methods and marked point processes for quarterly series. The lowest MAE and MSE values for each series type is highlighted in boldface.

		Stineman		KF-AA		KF-STS	
Stat.	Category	Bagged	Regular	Bagged	Regular	Bagged	Regular
Min. MAE	DEM	185.5	350.9	**181.9**	274	184.7	327.6
	FIN	**237.1**	380.2	**237.1**	315.9	238.2	330.5
	IND	293.7	740.9	**261.2**	451.7	288.7	612.4
	MAC	95.7	237.8	**86.1**	161.5	92.4	178.6
	MIC	323.7	730	**283.7**	434.9	308.5	505.9
Min. MSE	DEM	3.64E+05	1.26E+06	3.55E+05	7.87E+05	**3.41E+05**	1.01E+06
	FIN	2.13E+05	4.93E+05	**2.03E+05**	2.93E+05	2.12E+05	3.30E+05
	IND	2.42E+05	1.49E+06	**1.82E+05**	5.51E+05	2.33E+05	1.06E+06
	MAC	2.67E+04	2.19E+05	**2.18E+04**	8.60E+04	2.54E+04	1.20E+05
	MIC	4.95E+05	1.62E+06	**4.19E+05**	5.83E+05	4.72E+05	8.52E+05

		Linear		WMA-Exp	
Stat.	Category	Bagged	Regular	Bagged	Regular
Min. MAE	DEM	184.4	349.2	187.5	346.6
	FIN	236.4	374.8	239.2	340.3
	IND	294	736.6	296.1	704.7
	MAC	95.7	231.2	93.3	203.6
	MIC	323.7	711.6	314.8	598.4
Min. MSE	DEM	3.64E+05	1.26E+06	3.46E+05	1.11E+06
	FIN	2.13E+05	4.75E+05	2.14E+05	3.85E+05
	IND	2.43E+05	1.49E+06	2.51E+05	1.44E+06
	MAC	2.68E+04	2.13E+05	2.57E+04	1.71E+05
	MIC	4.95E+05	1.56E+06	4.93E+05	1.10E+06

Table 5
Overall averages of MSE and MAE values of bagged imputation for each combination of block bootstrap methods and marked point processes for quarterly series.

Stat	BBM	MPP	Stineman	KF-AA	KF-STS	Linear	WMA-Exp
MAE	CBB	AMS	364.3	358.9	364.3	364.4	364.8
	MBB	AMS	364.4	359.3	364.6	364.6	364.9
	NBB	AMS	**272.7**	**245.7**	**262.7**	**272.4**	**269.2**
	CBB	IVS	369.0	362.8	368.2	369.2	368.8
	MBB	IVS	376.4	370.7	375.3	376.4	375.9
	NBB	IVS	365.1	360.4	364.8	365.2	365.3
MSE	CBB	AMS	5.52E+05	5.38E+05	5.54E+05	5.52E+05	5.54E+05
	MBB	AMS	6.27E+05	6.11E+05	6.29E+05	6.28E+05	6.29E+05
	NBB	AMS	**3.55E+05**	**2.93E+05**	**3.33E+05**	**3.55E+05**	**3.49E+05**
	CBB	IVS	5.62E+05	5.46E+05	5.60E+05	5.63E+05	5.62E+05
	MBB	IVS	6.89E+05	6.73E+05	6.86E+05	6.89E+05	6.88E+05
	NBB	IVS	5.61E+05	5.47E+05	5.60E+05	5.62E+05	5.62E+05

C for the MCAR and MAR cases, respectively. Similar to yearly series, bagged version of non-overlapping block bootstrap with AMS is the most accurate one within the considered BBM-MPP combinations in quarterly series under both of the MCAR and MAR cases. However, for the quarterly series, IVS implementation with non-overlapping block bootstrap does not produce close results to AMS implementation. A possible reason of this is that the autocorrelation structure of the quarterly series gets more complicated by the existence of seasonality.

Histograms of block lengths that give minimum MAE values for Stineman implementation and Kalman filters with auto-arima for the quarterly series are presented in Fig. 12 and Fig. C6 of Appendix C for the MCAR and MAR cases, respectively. Because the results are similar for the rest of imputation methods, they are omitted here. Under the MCAR case, the block length of 4 gives the smallest error values for most of the quarterly series for all series types. Existence of a more complex serial dependence structure in the quarterly series due to the seasonality requires longer block lengths than those for the yearly series. For the MAR case, we observe from Fig. C6 that higher block lengths give promising results in addition to the block length of 4. A possible reason of this is the need for more information to accurately fill in the longer gaps created by the MAR mechanism.

3.2.3. Monthly series

For the comparison of MSE performance of the bagged and regular implementations of imputation methods for monthly series, we averaged the MSE values corresponding to bagged implementations over the combinations of block bootstrap methods and marked point processes. The results are given in Fig. 13 and Fig. B3 of Appendix B for the MCAR case, and Figs. C7 and C8 of Appendix C for the MAR case. Except the Kalman filters with auto-arima, the bagged algorithm has lower MSE values than the regular implementation under all BBM-MPP combinations for all the considered imputation methods for both of the MCAR and MAR cases. For the Kalman filters with auto-arima, bagged implementation with NBB-AMS gives lower MSE values than its regular implementation for all missingness rates (See Tables D3 and D6).

We have similar inferences for MAEs that are presented in Fig. B3 of Appendix B and Fig. C8 of Appendix C as those observed for

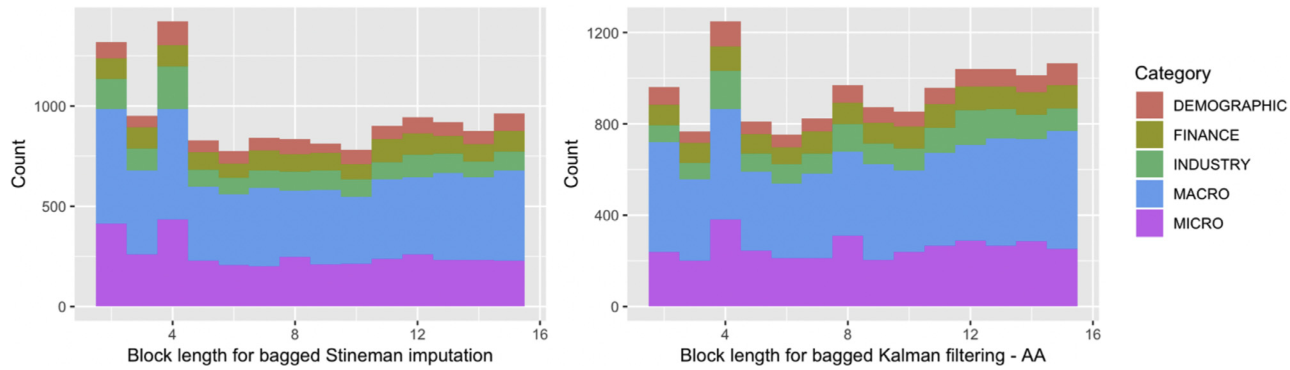

Fig. 12. Histogram of block lengths giving minimum MAE value for each series type for Stineman interpolation and Kalman filtering with auto-arima methods over all quarterly series.

Table 6
Average minimum MSE and MAE values for bagged implementation and average MAE and MSE values for regular implementation of imputation methods over block bootstrap methods and marked point processes for monthly series. The lowest MAE and MSE values for each series type is highlighted in boldface.

		Stineman		KF-AA		KF-STS	
Stat.	Category	Bagged	Regular	Bagged	Regular	Bagged	Regular
MAE	DEM	131.6	328.5	**131.1**	248.5	131.4	245.1
	FIN	293.7	578.5	**280.3**	322.4	291.4	410
	IND	380.5	708.8	**353.5**	449.9	365.4	520.8
	MAC	162.1	327.8	**151.8**	200.4	156.4	234.7
	MIC	852	1549.2	**742.0**	793.3	812.5	1077.8
	OTH	310.1	694.4	**303.6**	494.4	311.4	579.8
MSE	DEM	9.45E+04	5.96E+06	**9.33E+04**	2.75E+06	9.38E+04	2.94E+06
	FIN	5.54E+05	2.72E+06	**5.24E+05**	6.84E+05	5.49E+05	1.24E+06
	IND	5.01E+05	2.78E+06	**4.28E+05**	1.42E+06	4.52E+05	1.64E+06
	MAC	4.14E+05	1.07E+06	**2.86E+05**	3.24E+05	3.66E+05	4.88E+05
	MIC	1.73E+06	5.59E+06	**1.32E+06**	1.45E+06	1.56E+06	2.67E+06
	OTH	4.03E+05	2.00E+06	4.15E+05	8.42E+05	4.45E+05	1.22E+06

		Linear		WMA-Exp	
Stat.	Category	Bagged	Regular	Bagged	Regular
MAE	DEM	131.6	323	131.6	279
	FIN	293.7	568.8	292.9	494.7
	IND	379.7	696	374.4	622.7
	MAC	161.9	322	159.3	283.4
	MIC	850.6	1518	832.1	1288.8
	OTH	309.5	677	309.3	608.4
MSE	DEM	9.45E+04	5.79E+06	9.40E+04	3.81E+06
	FIN	5.53E+05	2.62E+06	5.51E+05	1.86E+06
	IND	4.98E+05	2.69E+06	4.78E+05	2.35E+06
	MAC	4.13E+05	1.03E+06	4.08E+05	7.45E+05
	MIC	1.72E+06	5.38E+06	1.65E+06	3.85E+06
	OTH	**3.99E+05**	1.91E+06	3.97E+05	1.47E+06

MSE values. In terms of MAE, the number of BBM-MPP combinations that give higher MAE for the bagged version than the regular version is higher for nearly all the imputation methods. However, specifically, bagged implementations with MBB-IVS and NBB-AMS have lower MAE values than their regular counterparts.

When we compare the bagged and regular versions of imputation methods, for all series types, we get lowest MAE and MSE values with the bagged algorithm (Table 6 for the MCAR case and Table C5 for the MAR case). In terms of both MAE and MSE, similar to yearly and monthly series, Kalman filters with auto-arima is the best performing imputation method under the bagged implementation for all series types but the “other” category. Weighted moving average with exponential weights has the lowest MSE for the series in the “other” category under the MCAR mechanism. When the longer periods of seasonality is introduced by the monthly series, overlapping BBM and IVS combinations give promising imputations in addition to the non-overlapping BBM and AMS combinations. With longer block lengths, overlapping BBM and IVS combinations include information about the serial correlation before and after the missing values (or block of missing values in the MAR case); and hence, they provide promising imputations for monthly series.

For monthly series, overall averages of MAE and MSE values for bagged and regular implementations for BBM-MPP combinations are given in Table 7 and Table C6 of Appendix C for the MCAR and MAR cases, respectively. More detailed results can be found in Tables D3 and D6 of Appendix D. In terms of MSE, CBB-AMS and MBB-IVS give lowest values for bagged implementation under the MCAR mechanism. For the MAR case, CBB-IVS produces the smallest MSE values. NBB-AMS provides the lowest MAE values for the bagged implementation under both of the MCAR and MAR cases. All these error values are significantly smaller than those

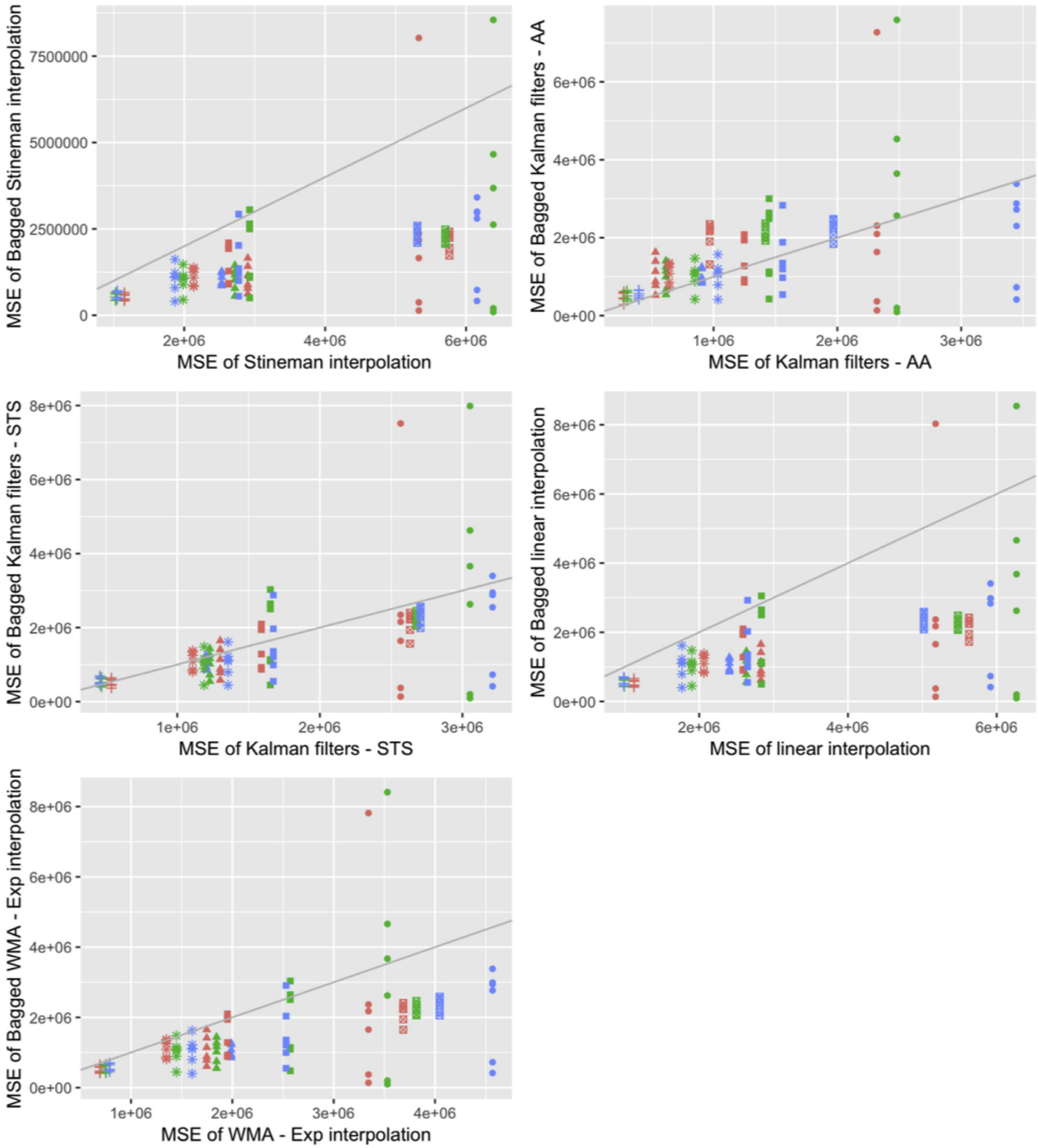

Fig. 13. Comparison of MSE values for the bagged and regular versions of interpolation methods for monthly series. The MSE values are averaged over BBM-MPP combinations for the bagged implementation. The lines on the panels show the 45-degrees regression line for comparison.

obtained with regular implementations (see Tables D3 and D6 of Appendix D).

Histograms of block lengths that give minimum MAE values for Stineman interpolation and Kalman filters with auto-arima are presented in Fig. 14 and Fig. C9 of Appendix C for the MCAR and MAR cases, respectively. Since the results for Stineman and linear interpolations and WMA-Exp are similar and those for the two versions of Kalman filters, they are omitted here and in Appendix C. The block length of 12 gives the smallest error values for monthly series for all series types under both of the missingness mechanisms. The required block length to get accurate imputations consistently increase along with the increasing frequency of the series. Getting minimum MAE values with the blocks of 12 months is consistent with the frequency of monthly series. As an exception, for Stineman and linear interpolations and WMA-Exp, the block length of 2 also provides promising imputations. Since WMA-Exp method considers a window of 4 data points prior to each missing observation, as seen here, it can provide promising imputation even with short block lengths.

Table 7
Overall averages of MSE and MAE values of the bagged imputation for each combination of block bootstrap methods and marked point processes for monthly series.

Stat.	BBM	MPP	Stineman	KF-AA	KF-STS	Linear	WMA-Exp
MAE	CBB	AMS	460.3	454.9	459.2	460.4	460.0
	MBB	AMS	478.9	473.6	478.4	479.1	478.9
	NBB	AMS	**426.3**	**396.6**	**414.0**	**425.6**	**420.9**
	CBB	IVS	503.1	497.5	501.9	503.1	502.6
	MBB	IVS	474.0	468.2	472.5	473.9	473.4
	NBB	IVS	496.8	492.5	496.4	496.9	496.7
MSE	CBB	AMS	**1.20E+06**	**1.16E+06**	**1.19E+06**	**1.20E+06**	**1.20E+06**
	MBB	AMS	1.52E+06	1.48E+06	1.51E+06	1.52E+06	1.51E+06
	NBB	AMS	1.92E+06	1.69E+06	1.81E+06	1.92E+06	1.88E+06
	CBB	IVS	1.76E+06	1.72E+06	1.75E+06	1.76E+06	1.76E+06
	MBB	IVS	1.25E+06	1.21E+06	1.24E+06	1.25E+06	1.25E+06
	NBB	IVS	1.55E+06	1.53E+06	1.55E+06	1.55E+06	1.55E+06

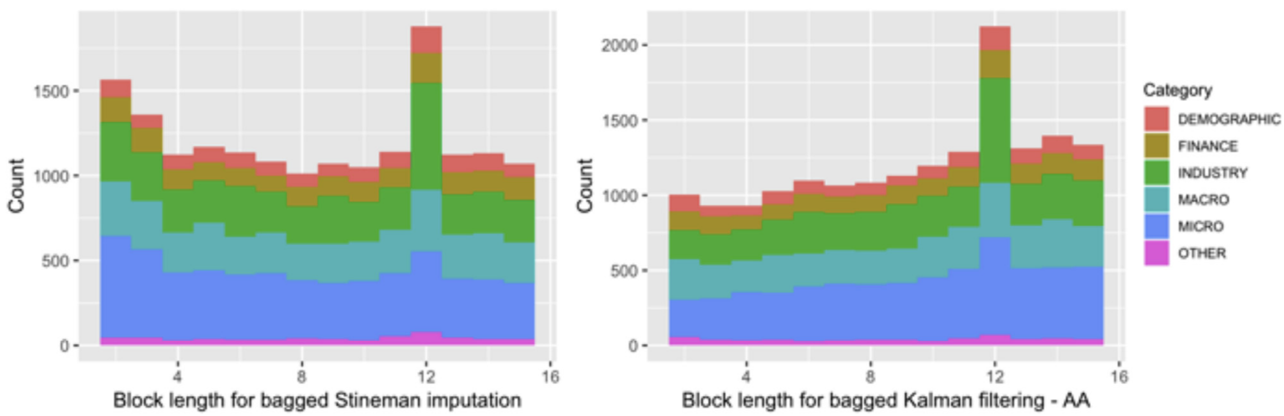

Fig. 14. Histogram of block lengths giving minimum MAE value for each series type for Stineman interpolation and Kalman filtering with auto-arima methods over all monthly series.

4. Conclusion and discussion

Missing observations in time series is a very common problem in nearly all applied fields where time dependent measurements are taken. The general approach to deal with the gaps arising from missingness is to impute the missing values using various missing value imputation methods. However, each imputation method has its own drawbacks. In this article, we approach the problem of missingness in time series data using bootstrap aggregation (bagging) context along with the combinations of block bootstrap methods and marked point processes. We propose a bagging algorithm to improve the accuracy of existing imputation methods. We utilise the marked point processes to preserve the missingness pattern during the bootstrap implementation and keep the missing values as is in bootstrap replications under the general bagging framework.

We benchmark the imputation accuracy of the proposed bagging algorithm with regular implementations of mainstream imputation methods Stineman and linear interpolation, Kalman filtering, and weighted moving average imputation over a real dataset including 2829 yearly, quarterly, and monthly demographic, financial, industrial, micro and macro economic time series of various lengths. This dataset originates from the M3 forecasting competition of forecasting practitioners. We consider two common missingness mechanisms, namely missing completely at random (MCAR) and missing at random (MAR), seen in practice. The following inferences are observed by our semi-Monte Carlo study where we generated missing values under the MCAR and MAR mechanisms across the series at three different rates:

- For all considered frequencies and missingness rates, the proposed bagging algorithm gives smaller mean square error and mean absolute error than the regular implementations of the considered imputation methods with at least one block bootstrap-marked point process combination under both of the MCAR and MAR cases. Therefore, we can conclude that the proposed algorithm improves on the regular implementation of imputation methods for time series data.
- The non-overlapping block bootstrap with amplitude modulated processes is the best performing approach in terms of imputation accuracy for all the considered frequencies and missingness mechanisms. For monthly series, which are relatively longer than yearly and quarterly series in the study, moving and circular block bootstrap methods with amplitude modulated processes also produce promising results. When the amplitude modulated process is combined with an overlapping block bootstrap method, if the block length and/or series is short, there is possibility of having blocks composed of missing observations, namely empty blocks (see Fig. 4(b)). However, when it is combined with a non-overlapping block bootstrap method, the possibility of having empty blocks is considerably less. Therefore, combination of amplitude modulated process and non-overlapping block bootstrap methods perform better. For longer series with higher frequency, overlapping block bootstrap and marked point process combinations include more information about the serial correlation before and after the missing values; and hence, they provide promising imputations at higher frequencies such as monthly series.
- Among the considered imputation methods, the proposed bagging algorithm with Kalman filters with auto-arima performs best for imputation of missing values in the series. The bagging provides the highest improvement in accuracy on Stineman imputation. This is in accordance with the nature of the bagging.

The effect of perturbing the series on the imputations is least in Kalman filters with auto-arima due to its auto-ARIMA feature. However, we still get a significant improvement in Kalman filters with auto-arima by the proposed bagging algorithm. This is due to capturing the serial correlation and seasonality structures of the series by the auto-arima feature of this method.

- In terms of the block length setting of the block bootstrap methods, for short yearly series, a block length of 2 gives the minimum error measures. When the frequency of the series is increased, serial correlation structure becomes more complicated; hence, the system needs more information to produce accurate imputations. This results in the need for longer block lengths for higher frequency series. By the increasing frequency and the length of the series, block lengths of 4 and 12 provide minimum errors for quarterly and monthly series, respectively. We suggest using higher block lengths for longer, high-frequency series.

In the overall sense, for the imputation of missing values in time series, we recommend bagging the Kalman filters with auto-arima by employing the proposed algorithm with the combination of non-overlapping block bootstrap and amplitude modulated processes.

It should be noted that computation time of the proposed approach increases as both the length of series and the number of bootstrap replications increase. In our simulations, average computation time was 0.48 ms per observation and bootstrap replication over the five considered imputation methods for the proposed approach and 0.55 ms per observation for the regular imputation methods on an iMac computer with Intel Core i7, 3.3 GHz processor. For example, implementation of our approach for a series of 100 observations takes 48 s with 1000 bootstrap replications on the average. Although this is considerably longer than the implementation of a regular imputation method, it is still a reasonable computation time with today's computers.

For the implementation of the proposed methodology in this article, the authors have prepared an R package called **baggedImpute**. The version 0.0.1 of the **baggedImpute** package is available on the GitHub platform (https://github.com/demirhanhaydar/baggedImpute) and it can be installed by the following R code chunk:

```
library(devtools)
install_github("demirhanhaydar/baggedImpute")
```

As a future study, implementation of a similar bagging algorithm with combinations of block bootstrap methods and marked point processes would be considered for multivariate irregular series.

Credit authorship contribution statement

Agung Andiojaya: Conceptualization, Methodology, Software, Writing - original draft, Visualization. **Haydar Demirhan:** Conceptualization, Methodology, Investigation, Validation, Visualization, Writing - original draft, Writing - review & editing.

Acknowledgements

The first author is supported by Indonesia Endowment Fund for Education (Lembaga Pengelola Dana Pendidikan - LPDP) scholarship with grant number 20160612 02 7311. The authors would like to thank three reviewers for detailed comments that considerably improved clarity and quality of the article.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.eswa.2019.03.044

References

Bashir, F., & Wei, H.-L. (2018). Handling missing data in multivariate time series using a vector autoregressive model-imputation (VAR-IM) algorithm. *Neurocomputing, 276*, 23–30.
Beck, M., & Bokde, N. (2017). *ImputeTestBench: Test bench for the comparison of imputation methods*. R package version 3.0.1.
Beretta, L., & Santaniello, A. (2016). Nearest neighbor imputation algorithms: A critical evaluation. *BMC Medical Informatics and Decision Making, 16*(3), 74.
Bermúdez, J. D., Corberán-Vallet, A., & Vercher, E. (2009). Forecasting time series with missing data using holt's model. *Journal of Statistical Planning and Inference, 139*(8), 2791–2799.
Bokde, N., Beck, M. W., Álvarez, F. M., & Kulat, K. (2018). A novel imputation methodology for time series based on pattern sequence forecasting. *Pattern Recognition Letters, 116*, 88–96.
Breiman, L. (1996a). Bagging predictors. *Machine Learning, 24*(2), 123–140.
Breiman, L. (1996b). Heuristics of instability and stabilization in model selection. *The Annals of statistics, 24*(6), 2350–2383.
Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *The Annals of Statistics*, 1171–1179.
Chatzis, S. P., Siakoulis, V., Petropoulos, A., Stavroulakis, E., & Vlachogiannakis, N. (2018). Forecasting stock market crisis events using deep and statistical machine learning techniques. *Expert Systems with Applications, 112*, 353–371.
Datta, D., & Du, W. (2012). Nonparametric HAC estimation for time series data with missing observations. *Technical Report 1060*. Board of Governors of the Federal Reserve System, International Finance Discussion Papers.
Demirhan, H., & Renwick, Z. (2018). Missing value imputation for short to mid-term horizontal solar irradiance data. *Applied Energy, 225*, 998–1012.
Eckner, A. (2012). A framework for the analysis of unevenly spaced time series data. Preprint. Available at: http://www.eckner.com/papers/unevenly_spaced_time_series_analysis.
Foroozand, H., & Weijs, S. V. (2017). Entropy ensemble filter: A modified bootstrap aggregating (bagging) procedure to improve efficiency in ensemble model simulation. *Entropy, 19*(10), 520.
Fung, C. (2006a). *Methods for the estimation of missing values in time series*. Western Australia: Edith Cowan University.
Fung, D. S. (2006b). *Methods for the estimation of missing values in time series. Western Australia: Edith Cowan University*.
Gómez, V., Maravall, A., & Peña, D. (1999). Missing observations in arima models: Skipping approach versus additive outlier approach. *Journal of Econometrics, 88*(2), 341–363.
Grewal, M., & Andrews, A. (2008). *Kalman filtering: Theory and practice using MATLAB* (3rd ed.). Wiley.
Hall, P. (1985). Resampling a coverage pattern. *Stochastic Processes and Their Applications, 20*(2), 231–246.
Hall, P., & Horowitz, J. L. (1993). *Corrections and blocking rules for the block bootstrap with dependent data*. University of Iowa, Department of Economics, Working Papers 93-11.
Hyndman, R., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: The state space approach*. Springer Science & Business Media.
Jain, A., & Kanhangad, V. (2018). Gender classification in smartphones using gait information. *Expert Systems with Applications, 93*, 257–266.
Jones, R. H. (1962). Spectral analysis with regularly missed observations. *The Annals of Mathematical Statistics*, 455–461.
Kumutha, V., & Palaniammal, S. (2013). An enhanced approach on handling missing values using bagging K-NN imputation. In *Computer communication and informatics (ICCCI), 2013 international conference on* (pp. 1–8). IEEE.
Kunsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 1217–1241.
Le, N. T., & Benjapolakul, W. (2018). A data imputation model in phasor measurement units based on bagged averaging of multiple linear regression. *IEEE Access, 6*, 39324–39333.
Lee, T.-H., & Yang, Y. (2006). Bagging binary and quantile predictors for time series. *Journal of Econometrics, 135*(1–2), 465–497.
Lee, T.-H., & Yang, Y. (2008). Chapter 13 bagging binary and quantile predictors for time series: Further issues. In *Forecasting in the presence of structural breaks and model uncertainty* (pp. 477–534). Emerald Group Publishing Limited.
Lepot, M., Aubin, J.-B., & Clemens, F. H. (2017). Interpolation in time series: An introductive overview of existing methods, their performance criteria and uncertainty assessment. *Water, 9*(10), 796.
Liang, G., Zhu, X., & Zhang, C. (2011). An empirical study of bagging predictors for imbalanced data with different levels of class distribution. In *Australasian joint conference on artificial intelligence* (pp. 213–222). Springer.
Lin, R., & Yin, G. (2016). Bootstrap aggregating continual reassessment method for dose finding in drug-combination trials. *The Annals of Applied Statistics, 10*(4), 2349–2376.
Little, R. J., & Rubin, D. B. (2014). *Statistical analysis with missing data*: Vol. 333. John Wiley & Sons.

Liu, A., Zhang, Y., Gehan, E., & Clarke, R. (2002). Block principal component analysis with application to gene microarray data classification. *Statistics in medicine, 21*(22), 3465–3474.

Liu, R. Y., & Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *In: R. lepage and l. billard, eds., exploring the limits of bootstrap* (pp. 225–248). New York: John Wiley.

Liu, Y., Ouyang, C., & Li, J. (2017). Ensemble method to joint inference for knowledge extraction. *Expert Systems with Applications, 83*, 114–121.

Louzada, F., Anacleto-Junior, O., Candolo, C., & Mazucheli, J. (2011). Poly-bagging predictors for classification modelling for credit scoring. *Expert Systems with Applications, 38*(10), 12717–12720.

Makridakis, S., & Hibon, M. (2000). The m3-competition: Results, conclusions and implications. *International Journal of Forecasting, 16*(4), 451–476.

Moritz, S., & Bartz-Beielstein, T. (2015). *imputets: Time series missing value imputation in R. R package version 0.4*.

Moritz, S., Sardá, A., Bartz-Beielstein, T., Zaefferer, M., & Stork, J. (2015). Comparison of different methods for univariate time series imputation in R. arXiv:1510.03924.

Niebuhr, T. (2017). Irregularly observed time series–some asymptotics and the block bootstrap. *Statistics, 51*(5), 1118–1131.

Parzen, E. (1963). On spectral analysis with missing observations and amplitude modulation. *Sankhyā: The Indian Journal of Statistics, Series A*, 383–392.

Politis, D. N., & Romano, J. P. (1991). A circular block-resampling procedure for stationary data. *Technical Report 370.*. Stanford University, Department of Statistics.

Ryan, K., & Giles, D. (1998). *Testing for unit roots with missing observations (department discussion papers no. 9802)*. Department of Economics, University of Victoria.

Shao, X. (2010). The dependent wild bootstrap. *Journal of the American Statistical Association, 105*(489), 218–235.

Simidjievski, N., Todorovski, L., & Džeroski, S. (2015). Predicting long-term population dynamics with bagging and boosting of process-based models. *Expert Systems with Applications, 42*(22), 8484–8496.

Stineman, R. W. (1980). A consistently well-behaved method of interpolation. *Creative Computing, 6*(7), 54–57.

Suzuki, T., & Ohkura, Y. (2016). Financial technical indicator based on chaotic bagging predictors for adaptive stock selection in japanese and american markets. *Physica A: Statistical Mechanics and its Applications, 442*, 50–66.

Zheng, X. (2018). Imputation-based HLA typing with SNPs in GWAS studies. In *HLA typing* (pp. 163–176). Springer.

Zheng, X., Shen, J., Cox, C., Wakefield, J. C., Ehm, M. G., Nelson, M. R., et al. (2014). HIBAG - HLA genotype imputation with attribute bagging. *The Pharmacogenomics Journal, 14*(2), 192.