Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v11n1p42

Forecasting financial short time series

By Alonso et al.

Published: 26 April 2018

Electronic Journal of Applied Statistical Analysis
Vol. 11, Issue 01, April 2018, 42-57
DOI: 10.1285/i20705948v11n1p42

Forecasting financial short time series

Andrés M. Alonso[a], Clara Simon de Blas[b], Ana Elizabeth García*[b], Mauricio Ciprián[c], Teresa Correas[c], Roberto Maestre[c], and Luis Peinado[c]

[a] *Universidad Carlos III de Madrid, Departament of Statistics and IFL*
[b] *Universidad Rey Juan Carlos, Statistics and Operation Research.*
[c] *BBVA Data and Analytics*

Published: 26 April 2018

In the present paper, we study the application of time series forecasting methods to massive datasets of financial short time series. In our example, the time series arise from analyzing monthly expenses and incomings personal financial records. Unlike from traditional time series forecasting applications, we work with series of very short depth (as short as 24 data points), which does not allow us to use classical exponential smoothing methods. However, this shortcoming is compensated by the size of our dataset: millions of time series. This allows us to tackle the problem of time series prediction from a pattern recognition perspective. Specifically, we propose a method for short time series prediction based on time series clustering and distance-based regression. We experimentally show that this strategy leads to improved accuracy compared to exponential smoothing methods. In addition, we describe the underlying big data platform developed to carry out the efficient forecasting, since we perform millions of item comparisons in near real-time.

Keywords: Financial time series, Big data, Forecasting, Conditional mean, Holt Winter, Clustering.

*Corresponding author:anaelizabeth.garcia@urjc.es

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

1 Introduction

This study deals with the forecasting of monthly outgoings and incomings personal financial records. Here, the input is a time series where each point represents an expense or incoming in a particular financial category, aggregated over a predefined time interval. We aim to output the most likely value of that expense or income for the subsequent time interval. Examples of categories include fine-grained expense type indicators such as “utility bill”, “salary” or “ATM withdrawal”. The chosen unit of aggregation in the month, as this is a natural unit of personal financial planning. One potential application of this functionality is that of personal finance management APPS. In this example, customers of a financial institution would benefit from digital tools to anticipate future expenses and hence better plan their budgets.

There are many statistical techniques available for time series forecasting that can be used to deal with the aforementioned problem. De Gooijer and Hyndman (2006) provide a survey paper on the topic. However, particular time series display some unique properties.

Firstly, since the selected time unit is the month, while the history of the series spans a period 2 years, our time series contain 24 points, which can be considered relatively short. Second, the distribution of expenses in categories follows a long tail distribution where a few categories concentrate the most common expenses and a high number of categories appear occasionally. Thus, the dataset contains many sparse series. These two properties represent a challenge for time series forecasting since methods such as exponential smoothing typically work with longer series.

Finally, a third outstanding property of our dataset is its size. As our dataset contains thousands of customers and hundreds of categories, we work with millions of time series. We see this as an advantage rather than as a limitation, as it allows us to pose the problem of time series prediction as a pattern recognition. We can interpret series as “patterns” and use similar patterns to inform the inference of the following value of the series.

In this paper, we compare a classical times series forecasting approach against a methodology based on pattern recognition for the problem of short time series prediction. More specifically, we study the application of the well-known Holt-Winters exponential smoothing method, and of an approach based on nearest-neighbor regression. In the latter, the forecast is considered a missing variable to be inferred, and is extrapolated using the (known) values of that variable in the most similar series in the training set. Experiments show that, in a situation with a large number of short time series, the pattern-based approach outperforms that of the classical time series. A possible explanation is that exponential smoothing methods cast a prediction by only using intra-series information. Therefore, using a mere 24 data points may be insufficient. In contrast, nearest-neighbor exploits inter-series information and propagates prediction using similarity patterns between pairs of millions of time series. The underlying assumption that similar series tend to exhibit a similar value of the variable make predictions seems to be validated by our experiments. As a consequence of our study, we have obtained a robust method for predicting expenses and incomings a real-world financial application. For comprehensive reviews on time series forecasting, the reader is referred to De Gooijer and Hyndman (2006). Exponential smoothing is discussed in depth in Hyndman et al. (2008).

The remainder of the paper is organized as follows: a brief introduction of the problem is given in section 1. Section 2 describes the time series dataset. The methodology and procedures applied can be found in section 3. Finally, the paper is concluded in section 4 providing the main results.

2 Dataset and notation

The dataset used in this paper was built from the BBVA account transaction records. As a financial institution, BBVA maintains an exhaustive record of all account transactions executed by its clients. Each transaction record is composed of a set of fields including the contract id of the account, transaction type, transaction date and transaction amount. Transaction type is a code that describes the different actions that clients can carry out with their accounts. There are more than 1,000 transaction types such as ATM disposal, Salary income, Utilities, Money Transfer, Credit card bill, among others.

Using this data, we generated a dataset using the transactions executed from July 2013 to July 2015 by 5,000 randomly selected contracts. Their transactions were aggregated month by month in each one of the top 20 most common transaction types among BBVA clients. The result is a set of 32,479 monthly time series, each one representing accumulated income/outgoing for an account and transaction type during the course of 25 consecutive months. Therefore, our dataset is composed of the following fields:

- Contract id: id of the account where the transaction is executed. This contract id has been encrypted in order to preserve the privacy of the account holder.
- Transaction type: Code that represents the type of transaction executed on the account. This dataset contains 20 different transaction types, table 1 shows the distribution of transaction types in the dataset.
- The following 25 fields $t_0, t_1, t_2, ..., t_{24}$ represent the accumulated income/outgoing time series for the account id and transaction type month by month from July 2013 to July 2015.

The diversity of people's economic behavior represented in the dataset, along with the reduced size of the time series (24 values + one value to predict), make it an interesting and challenging dataset for forecasting.

3 Prediction Procedures

This section highlights some improvements to the existing transaction value forecasting methods used by the bank. The current methods for bank account transaction modelling for monthly time series are:

Transaction Type	# Time Series in Dataset	% of Dataset
ATM Disposal	3,500	10.78%
Direct debit	3,384	10.42%
Bank Transfer	3,150	9.70%
Pay checks	2,056	6.33%
Debit card payment in supermarket	1,729	5.32%
Credit card bill	1,590	4.90%
Retirement pay	1,568	4.83%
Deposit in office	1,551	4.78%
Disposal in office	1,410	4.34%
Debit card payment in fashion stores	1,363	4.20%
Debit card payment in gas stations	1,294	3.98%
Telephone bill	1,224	3.77%
Debit card payment in superstores	1,197	3.69%
Utilities payment	1,186	3.65%
Debit card payment in furniture stores	1,097	3.38%
Utilities payment (2)	1,088	3.35%
Telephone bill (2)	1,081	3.33%
Card payment fee	1,022	3.15%
Insurance bill	995	3.06%
Loan repayment	994	3.06%
Total	**32,479**	**100%**

Table 1: Transaction types in dataset

- Procedure based on the time series average. Only applied to series with small transaction frequency dispersion.
- Procedure based on cluster averages.

First, we will present a selection criterion for admissible time series. Thereafter, we consider two years monthly time series. We consider three alternative criteria:

A. We use the observed times between transactions, that is, if we have a time series $(X_1, X_2, \ldots, X_{24})$ and we observe non null values in the months $(i_1, i_2, \ldots, i_k)$ then we have the following observed $k-1$ times between the transactions $i_2 - i_1$, $i_3 - i_2$, $\ldots$, $i_k - i_{k-1}$. Hence, the procedure will provide a forecast only for series having a coefficient of variation less than 0.2 for those times.

   This criterion has the disadvantage of not detecting anomalous situations where the series has few transactions at the beginning or at the end of the series. In this criterion, A, we use only the observed times between transactions.

B. We add two estimated times between transactions to the previous criterion A, one at the beginning of the series, i_1, and other at the end, $25 - i_k$. This is equivalent to assuming that there was an account transaction at month zero and there will be a transaction at month 25. Therefore, the procedure will provide a forecast only for series having a coefficient of variation less than 0.2 for these times.

   We add a zero at the beginning and a *25* at the end in order to estimate the first and the last time of a transaction, respectively. Therefore, we can have at most 23 observed times and two estimated times. This criterion solves the aforementioned disadvantage since a series with few transactions at the beginning would have a high dispersion since $25 - i_k$ would be high and a series with few transactions at the end would have also a high dispersion since i_1 would be also high. On the other hand, the number of transactions of the series included according to this criterion are 8 (or 7), 12 (or 11) and 24 (or 23), in other words, series with quarterly, bimonthly or monthly transactions.

C. An intermediate alternative is using criteria A's observed times and the first estimated time. Though this procedure solves the anomalous situations when transactions are concentrated at the end of the series it fails when transactions are concentrated at the beginning of the time series.

   We add a zero at the beginning in order to estimate the first time for a transaction. This means that, we can have at most 23 observed times and one estimated time.

The results of applying these three criteria to the 32479 series in the dataset reveals the following conclusions:

- 11302, 9974 and 10878 series satisfy criteria A, B and C, respectively.
- The series that satisfy criteria B also satisfy criteria C.

- The series that satisfy criteria C or B do not necessarily satisfy criteria A.

Given the previous results, it is difficult to compare the prediction results when applying criterion A versus B/C. For case B versus C, it appears that the 904 additional series do not deteriorate the prediction performance. For the three criteria we will show the results of improvement proposals in the procedure of prediction based on the mean, although in some cases we focus on criteria C.

The rest of the section is organized in two parts: the first one deals with procedures based on the average, an AR-benchmark and three possible improvements (conditional mean; conditional mean with selected lags and Holt-Winters approach, Holt, 2004), the second part is based on neighbours–*cluster* and two possible improvements (weighted mean and weighted mean with blocks).

3.1 Procedures based on the time series average

3.1.1 Procedure based on the conditional mean (CM)

The idea of this procedure is to use a kernel estimator where in the mean computation observations are weighted by factors that depend on the distance between the last available observation and the previous observations. The procedure can be described as follows Bosq (2012):

1. Given a time series $(X_1, X_2, \ldots, X_{24})$ we construct pairs of values (X_1, X_2), (X_2, X_3), $\ldots$, (X_{23}, X_{24}).
2. We compute the difference between X_{24} and the initial values of the pairs constructed in step 1.
3. Weights are proportional to $K\ (X_i - X_{24})$, that is to the Gaussian kernel K evaluated in the previous differences. [1]
4. Finally, the forecast is calculated using the expression:

$$\widehat{X}_{25} = \sum_{j=1}^{23} W_j X_{j+1},$$

where

$$W_j = \frac{K(X_j - X_{24})}{\sum_{j=1}^{23} K(X_j - X_{24})}.$$

This procedure can be implemented considering the latest observations for obtaining the weights. This can be advantageous in series that have a bimonthly or quarterly behavior.

[1] Note that the kernel used is a symmetric function with respect to zero, i.e., $K(x) = K(-x)$. Similarly, it can be interpreted as weights proportional to $K(x)$.

3.1.2 Procedure based on the conditional mean with lag selection (CM-lag)

A more general implementation of the conditional mean procedure allows us to include a simple criterion to select the number of observations to condition the prediction or number of lags used for forecasting. As a preliminary step the autocorrelation function is calculated. The number of lags will be selected as the order of the greater autocorrelation in absolute value. The underlying idea is that, for example, in a series of quarterly transactions we can observe a significant correlation of order three greater than the two previous ones. In such cases, the prediction is made conditional to the value observed in (X_{22}, X_{23}, X_{24}).

Table 2 presents the results of the simulations of previous procedures and the procedure based on the mean for the three inclusion criteria. Error1 and Error2, Error3 terms correspond to the absolute errors of the procedure based on the mean, CM and CM-Lag, respectively. Also in the table, in order to have a benchmark, we will consider the forecasting results using autoregressive models estimated by the Yule-Walker procedure. The order of the autoregressive model was selected by the Akaike Information Criterion (AIC, see Akaike, 1969) and the maximum considered order was twelve. Similar results were obtained by using the Bayesian Information Criterion (BIC, see Schwarz et al., 1978).

In the three criteria (A, B and C), the absolute errors statistics obtained with the AR-benchmark are better than the results with the mean procedure, our proposals, CM and CM-Lag, outperforms both the mean procedure and the benchmark. In the three criteria is observed that the CM and CM-lag procedures improve substantially the mean procedure in all the considered statistics with reductions nearby or top to the 50% in the median of the absolute errors. The differences between CM and CM-Lag are not noticeable and there is even some deterioration in the third quartile of the absolute errors.

On the other hand, the previous results point out the need for an ex-ante criterion to determine if we can expect a reasonable low forecast error for a given time series. A preliminary graphical analysis reveals that there are some atypical series among the series with the same code. There are at least two possible types of atypical series:

1. Due to level shift in the complete period of observation.

2. Due to some atypical behavior at some point of the period of observation.

Some examples of both types appear in time series corresponding to code 80. The first type of atypical series can be dealt with by standardizing the series (see Figure 1) but the second type could be masked by this procedure. That is, series that are far away from the data cloud could be inside the cloud of standardized points. We labeled the atypical series in a multivariate approach and we classify a time series as an outlier if it has less than N neighbors. We have considered N=100 neighbors.

As we can see in Table 3, the exclusion of 396 atypical series detected by the neighbors procedure improves the absolute errors statistics, and especially the higher errors attached in the third quartile and the maximum error. The results are shown with the

Criteria A					
	min	1st Q	median	3rd Q	max
Error1	0	25.6667	80.3625	225.2083	4.0644e+05
Error2	0	4.7260	32.3756	147.3149	3.0750e+05
Error3	0	3.3834	29.6796	151.4523	3.0750e+05
AR benchmark	0	10.6880	63.9570	202.0800	4.0644e+05
Criteria B					
	min	1st Q	median	3rd Q	max
Error1	0	33.7983	110.6183	270.9983	4.0644e+05
Error2	0	13.0026	60.6834	228.7785	3.0750e+05
Error3	0	10.6190	59.9367	238.2817	3.0750e+05
AR benchmark	0	26.6730	103.2500	263.1400	4.0644e+05
Criteria C					
	min	1st Q	median	3rd Q	max
Error1	0	28.9667	103.2500	265.6583	4.0644e+05
Error2	0	9.0909	49.7885	205.9539	3.0750e+05
Error3	0	6.2268	47.9425	214.4715	3.0750e+05
AR benchmark	0	20.0220	93.3570	248.0000	4.0644e+05

Table 2: Absolute prediction error statistics for series satisfying A, B and C. The terms Error1, Error2, Error3 and AR-benchmark correspond to the absolute errors of the procedure based on the mean, CM, CM-Lag and AR models, respectively.

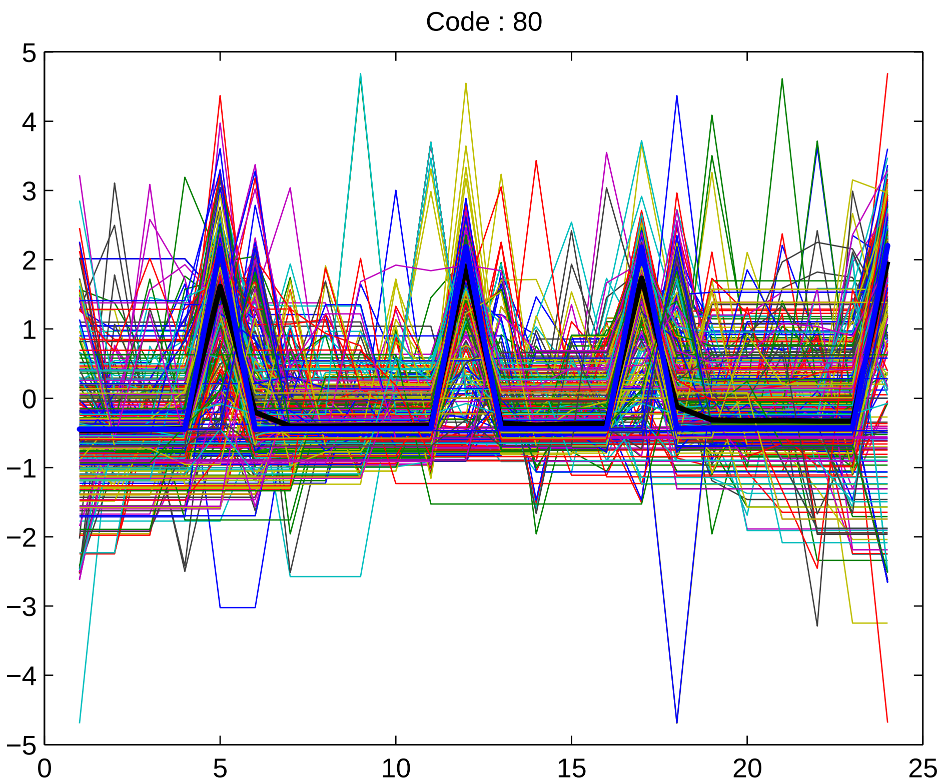

Figure 1: Time series for all customers with account transactions in code 80.

Criteria C					
	min	1st Q	median	3rd Q	max
Error1	0	27.5400	100.8100	247.9200	6.0636e+04
Error2	0	8.3238	46.2410	189.6000	6.0596e+04
Error3	0	5.5612	44.5340	196.9300	6.0596e+04
AR benchmark	0	18.7500	87.7270	233.5300	6.0636e+04

Table 3: Absolute prediction error statistics for series satisfying C and excluding atypical series. The terms Error1, Error2, Error3 and AR-benchmark correspond to the absolute errors of the procedure based on the mean, CM, CM-Lag and AR models, respectively.

set of series that satisfy criterion C but the conclusion is extensible to the rest of the criteria. In table 3, we also include the results with the AR-benchmark. Again, the AR-benchmark is superior to the mean procedure but it is outperformed by CM and CM-Lag. These results suggest that CM and CM-Lag take advantage of the flexibility of kernel estimators to approximate non-linear relationships.

3.1.3 Holt Winters procedure (HW)

Holt (2004) and Winters (1960) extended Holt's method to capture seasonality. The Holt-Winters seasonal method comprises the forecast equation and three smoothing equations; one for the level, one for trend, and one for the seasonal component with smoothing parameters. We have implemented the Holt-Winters seasonal method in the following ways:

(a) HW1: Implements the Holt-Winters procedure with trend and seasonality, gives as output the absolute mean values errors and the forecast a month view, for annual series. It performs the optimal selection of the parameters alpha, beta and gamma within the interval [0.1, 1] with increments of 0.1, according to the criterion of minimizing absolute mean errors of all series.

(b) HW2: Implements the Holt-Winters procedure without trend and seasonality

(c) HW3: Implements the Holt-Winters procedure with trend

(d) HW4: Implements exponential smoothing procedure, i.e. without trend and seasonality. The simple exponential smoothing is a compromise between predictions by the latest data and predicted by the mean; It makes a prediction based on a weighted average of the current values and the last value. When calculating this mean, a higher weight is assigned to the most recent observation, lowering gradually for the preceding observations.

Criteria C						
	min	1st Q	median	3rd Q	max	NaN
AE1	0.0069	47.142	140.73	348.36	7.2995e+05	2102
AE2	0	8.7744	60.5065	255.7845	9.4932e+15	2102
AE3	0	17.2804	121.4202	572.0091	7.5852e+05	-
AE4	0	13.1245	106.3291	571.001	6.5081e+05	-
AE5	0	3.1140	52.5458	256.5759	7.5852e+05	-

Table 4: Absolute prediction error statistics for series satisfying C. The terms AE1, AE2, AE3, AE4 and AE5 correspond to the absolute prediction errors of procedures HW1, HW2, HW3, HW4 and HW5, respectively.

(e) HW5: Calculates for all the series that satisfy criterion C, the optimal parameters and selects the procedure of prediction in the four previous variants based on the absolute error prediction of the last observation available in the sample

Table 4 shows the results for all time series in the dataset, when implementing the above procedures based on the mean for the criterion inclusion C. The AE1, AE2, AE3, AE4 and AE5 terms correspond to the absolute errors prediction of the HW1 procedure with seasonality and trend, HW2 seasonality and trend, HW3 no seasonality and trend, HW4 exponential smoothing and HW5 that selects the procedure with the lowest absolute prediction error of the last observation.

HW1 and HW2 procedures that consider seasonality are not applicable to all the series, particularly when there is a high frequency of zeros (in this data set we can detect 2102 series). This inconvenience is not presented in HW3 and HW4 procedures, and therefore would be eligible procedures for such cases.

When comparing the above results with Error1 in Table 2, we note that HW5 outperforms the procedure based on the mean of the three quartiles but not in the maximum error. On the other hand, HW5 improves CM and CM-Lag in the first quartile but not in the rest of the statistics.

Table 5 shows the results when we exclude the 396 (less than 400) atypical series detected by the neighbors procedure. As expected, the exclusion of these atypical series improves the absolute errors statistics.

An important element of the HW5 procedure is the selection of the Holt-Winters model used, that is, which criteria to select among the four variants HW1 - HW4. In the present study we have used the minimum of $\left|X_{24} - \widehat{X}_{24}\right|$. It has also been considered the lowest mean absolute error (MAE) in-sample as a selection criteria, however, the results only improve the first and second quartiles to Error1. On the other hand, the good results of the conditional procedures CM and CM-Lag suggest that the selection criteria should be conditional.

Criteria C						
	min	1st Q	median	3rd Q	max	Nans
AE1	0.0069	44.8823	135.1015	329.7792	3.2032e+05	2032
AE2	0	7.9617	56.7867	233.2105	9.4932e+15	2032
AE3	0	16.3420	116.5342	571.0001	6.0119e+04	
AE4	0	12.1082	102.6322	562.8992	5.9851e+04	
AE5	0	2.9407	49.9687	234.5000	5.9449e+04	

Table 5: Absolute prediction error statistics for series satisfying C and excluding atypical series. The AE1, AE2, AE3, AE4 and AE5 terms correspond to the absolute prediction errors of procedures HW1, HW2, HW3, HW4 and HW5, respectively.

3.2 Procedure based on the neighbors (cluster) mean

This procedure can be considered within the family of conditional prediction procedures, in the sense that we are trying to detect series with a pattern in the observations $(X_1, X_2, \ldots, X_{12})$ similar to the pattern of the series to be predicted, Y, in the observations $(Y_{13}, Y_{14}, ..., Y_{24})$. The procedure can be described as:

1. Standardized series which belong to the same code. Denote them by ($\mathbf{Z}_1$, $\mathbf{Z}_2$, $\ldots, \mathbf{Z}_N$).

2. Let $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_{24})$ be the time series to be predicted and let $\mathbf{Z} = (Z_1, Z_2, ..., Z_{24})$ be its corresponding standardization.

3. We compute the Euclidean distances between the observations (Z_{13}, Z_{14}, $\ldots, Z_{24}$) of the series $\mathbf{Z}$ and the observations ($Z_i(1)$, $Z_i(2)$, $\ldots, Z_i(12)$ of the series $\mathbf{Z}_i$ with $i = 1, 2, \ldots, N$.

4. We obtain the k nearest neighbors with respect to the distances calculated in the previous step. Let's denote the set of neighbors by $K_z = \{i_1, i_2, \ldots, i_k\}$

5. We compute the forecast as:

$$\widehat{X}_{25} = \overline{Y} + \overline{Z}_{25} \times S_Y,$$

where $\overline{Y}$ and S_Y are the mean and the standard deviation of the series to be forecast, respectively, and

$$\widehat{Z}_{25} = \frac{1}{k} \sum_{i \in K_z} Z_i(13)$$

i.e., the mean values of the thirteen months among the Y's neighbors.

All time series					
	min	1st Q	median	3rd Q	max
Clustererror1	0	20.136	59.738	203.120	3.8207e+05
KernelClusterError1	0	16.843	55.941	190.530	3.6426e+05
Criteria C					
	min	1st Q	median	3rd Q	max
Clustererror2	0	14.171	67.810	248.590	4.0719e+05
KernelClusterError2	0	10.756	61.863	235.410	4.0197e+05
Criteria C					
	min	1st Q	median	3rd Q	max
Clustererror3	0	13.850	64.914	238.040	3.8207e+05
KernelClusterError3	0	10.120	60.022	225.080	3.6426e+05

Table 6: Absolute prediction errors statistics for all series and for series satisfying criterion C. The terms ClusterError and KernelClusterError correspond to the absolute prediction errors of the cluster/neighbors mean and CMW mean, respectively.

3.2.1 A proposal for improvement based on the weighted mean (CMW).

In the procedure based on the mean of the time series we noted that the weighting type kernel improved the prediction results significantly. The first improvement is therefore to consider

$$\widehat{Z}_{25} = \frac{1}{k} \sum_{i \in K_z} W(i) Z_i(13),$$

where

$$W(i) = \frac{K\left(\|Z_i(1:12) - Z(13:24)\|\right)}{\sum_{i \in K_z} K\left(\|Z_i(1:12) - Z(13:24)\|\right)}.$$

Other distances can be considered besides the Euclidean in both the original implementation and in the proposed CMW. The results using the distance of Mahalanobis are similar to those obtained with the Euclidean distance.

Table 6 shows the results when implementing the two above procedures and the one based on the mean for all series and for those that satisfy the inclusion criterion C. The terms ClusterError1 and KernelClusterError1 correspond to the absolute errors prediction procedure based on the cluster/neighbors and CMW mean, respectively. The terms ClusterError2 and KernelClusterError2 are similar but only use the series that satisfy the criterion C in both the calculation and the prediction. While the terms ClusterError3 and KernelClusterError3 refer to the series that satisfy the criterion C but using all series in the same code for prediction.

All the series					
	min	1st Q	median	3rd Q	max
KernelClusterError1	0	13.888	51.527	188.930	3.4391e+05
Criteria C					
	min	1st Q	median	3rd Q	max
KernelClusterError2	0	9.279	55.771	218.250	3.822e+05
Criteria C					
	min	1st Q	median	3rd Q	max
KernelClusterError3	0	9.275	56.812	216.740	3.4391e+05

Table 7: Absolute prediction errors statistics for all series and for series satisfying criterion C and excluding atypical series. The terms KernelClusterError correspond to the absolute prediction errors of the CMW^2 mean.

Both cluster procedures improve the forecast obtained with the time series mean. The CMW procedure improves the procedure based on the neighbors mean both in the set of all series as in the subset of the series that satisfy the criterion C. We observe a slight improvement when considering all the sets of the same code for those time series that satisfy the criteria C.

3.2.2 A proposal for improvement based on the weighted mean by blocks ($\mathbf{CMW^2}$).

The positive results obtained with the CM procedure where time series blocks are used, lead us to consider other series blocks as possible neighbors. The search of neighbors or similar patterns will be considered in the sets of vectors:

$$\begin{array}{cccc} (X_1, & X_2, & \ldots, & X_{12}) \\ (X_2, & X_3, & \ldots, & X_{13}) \\ \vdots & \vdots & \vdots & \vdots \\ (X_{11}, & X_{12} & \ldots, & X_{23}) \end{array} .$$

This idea can be interpreted as a rigid time-warping where contractions or expansions of the temporary index are allowed. On the other hand, a more exhaustive use of all available information is made.

Table 7 shows the results of the CMW^2 procedure for all series and for those that satisfy the criteria C.

As we can observe the incorporation of the blocks in CMW^2 with respect to CMW, improves forecasting results in all the considered statistics. The procedures based on the neighbors (cluster) mean and the CMW, CMW^2 procedures used two parameters that should be chosen: (i) the number of neighbors to calculate the means (it was fixed to

all the series						
	min	1st Q	median	3rd Q	max	NaNs
ClusterError1	0	19.396	56.658	187.890	6.1222e+04	1163
KernelClusterError1	0	16.277	53.699	177.780	6.1023e+04	1163
Criteria C						
	min	1st Q	median	3rd Q	max	NaNs
ClusterError2	0	13.408	62.888	230.630	6.1641e+04	396
KernelClusterError2	0	10.102	57.572	217.100	6.0501e+04	396
Criteria C						
	min	1st Q	median	3rd Q	max	NaNs
ClusterError3	0	13.118	59.967	221.400	6.1222e+04	396
KernelClusterError3	0	9.463	55.840	208.670	6.1023e+04	396

Table 8: Absolute prediction errors statistics for all series and for series satisfying criterion C. The terms ClusterError and KernelClusterError correspond to the absolute prediction errors of the cluster/neighbors mean and CMW mean, respectively.

100 neighbors) and (ii) the size of the block to search for similar patterns/neighbors (it was fixed to 12 months).

On the other hand, as in the procedures based on the mean of the series, it is interesting to consider the exclusion of series that have an atypical behavior in the sense of having a "small" number of neighbors. Table 8 includes this option and presents the corresponding errors. As we expected, the exclusion of 1163 atypical series in the complete data set (and 396 atypical series in the set that satisfies criterion C leads to improvements in the forecast results.

4 Conclusions

In this paper we have selected two forecasting methods that are susceptible of improvement with techniques of easy implementation that are scalable to Big data sets. For each of the procedures we have proposed two or more alternatives for improvement. It has been illustrated that they are effective in the data set forecast. The best results are obtained with the CM and CMW^2 procedures. Some aspects that could be subject of future research are: Development of an ex-ante selection series procedure whose prediction has a correct behavior (in this way, we only offer the prediction services to the contracts that satisfy these selection criteria); the combination of the predictions of the best methods, CM with an univariate analysis (horizontal/intra-series) and CMW^2 with a multivariate analysis (vertical/inter-series). Finally, other definitions of neighborhood can be considered, and the combination with regression techniques may result in further improvements of the forecasting procedure.

Acknowledgement

The first and third authors acknowledge the support of Grant ECO2015-66593-P of MINECO/FEDER/UE.

References

Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the institute of Statistical Mathematics*, 21(1):243–247.

Bosq, D. (2012). *Nonparametric statistics for stochastic processes: estimation and prediction*, volume 110. Springer Science & Business Media.

De Gooijer, J. G. and Hyndman, R. J. (2006). 25 years of time series forecasting. *International journal of forecasting*, 22(3):443–473.

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International journal of forecasting*, 20(1):5–10.

Hyndman, R. J., Koehler, A. B., Ord, J. K., and Snyder, R. D. (2008). *Forecasting with exponential smoothing: the state space approach*. Springer Verlag.

Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management science*, 6(3):324–342.