Modelling and forecasting based on recurrent pseudoinverse matrices

Christos K. Filelis - Papadopoulos[1][0000−0002−6591−970X], Panagiotis E. Kyziropoulos[1][0000−0003−1882−342X], John P. Morrison[1][0000−0002−0134−6025], and Philip O'Reilly[2][0000−0001−9286−2107]

[1] University College Cork, Department of Computer Science, Western Gateway Building, Cork, Ireland
`christos.papadopoulos-filelis@cs.ucc.ie,panagiotis.kyziropoulos@ucc.ie,`
`j.morrison@cs.ucc.ie`
[2] University College Cork, Cork University Business School, O'Rahilly Building, Cork, Ireland
`philip.oreilly@ucc.ie`

Abstract. Time series modelling and forecasting techniques have a wide spectrum of applications in several fields including economics, finance, engineering and computer science. Most available modelling and forecasting techniques are applicable to a specific underlying phenomenon and its properties and lack generality of application, while more general forecasting techniques require substantial computational time for training and application. Herewith, we present a general modelling framework based on a recursive Schur - complement technique, that utilizes a set of basis functions, either linear or non-linear, to form a model for a general time series. The basis functions need not be orthogonal and their number is determined adaptively based on fitting accuracy. Moreover, no assumptions are required for the input data. The coefficients for the basis functions are computed using a recursive pseudoinverse matrix, thus they can be recomputed for different input data. The case of sinusoidal basis functions is presented. Discussions around stability of the resulting model and choice of basis functions is also provided. Numerical results depicting the applicability and effectiveness of the proposed technique are given.

Keywords: Forecasting · Pseudoinverse matrix · Modelling.

1 Introduction

General time series modelling and forecasting has become an essential tool in several scientific fields and business sectors, spanning from physics and engineering to workforce prediction and finance. Several methods have been proposed in the literature for modelling and forecasting time series based on either statistical or Machine Learning techniques. The most notable methods in the statistical methods class include Exponential Smoothing (ES), Auto-Regressive Integrated Moving Average (ARIMA), State Space and Structural models, Kalman filter,

Nonlinear models and Generalized Auto-Regressive Conditionally Heteroscedastic (GARCH) models.

Exponential Smoothing (ES) was proposed in the 1960s by Brown, Holt and Winters [3, 11, 27], has been used extensively along with its variants due to its simplicity and adaptability to various scientific fields [14, 25]. Auto-Regressive (AR), Auto-Regressive Moving Average (ARMA) and Auto-Regressive Integrated Moving Average (ARIMA) models were first proposed by Box and Jenkins [2] in the 1970s. Estimation techniques for the parameters of such methods have been discussed in [2]. Several univariate time series forecasting variants of such models have been proposed in the literature including ARARMA [18], Vector ARIMA (VARIMA) [21, 23], Automatic univariate ARIMA type models [24] and seasonal approaches such as STL or $X - 11$ [4, 5]. In the 1980s, state space models have been also used including Dynamic Linear Models (DLM) [9] and Basic Structural Models (BSM) [10] along with several variants [20]. In the quantitative finance domain Generalized Auto-Regressive Conditionally Heteroscedastic (GARCH) models [6] have been used. Nonlinear variants that handle asymmetric volatility have been also presented and studied [1, 17].

Research around non-linear modelling techniques has also been conducted however not in such extent. This is primarily because of their increased complexity and the lack of closed form formulas. In this field notable contributions include the work of Wiener [26] and others [19]. Detailed overview on the advances of time series modelling and forecasting techniques over the last decades are given in [8] and references inside. Moreover, large scale benchmarking of popular modelling and forecasting techniques are given in [15].

Despite the popularity and wide use of the aforementioned techniques, they are mostly dedicated tools tuned specifically to a use case relying on restrictive assumptions for the form and nature of the data. General models such as Artificial Neural Networks or Support Vector Machines handle those limitations but they require substantial amounts of data, increased training and retraining times, as well as substantial tuning of the large number of hyperparameters. A different approach is followed by Fast Orthogonal Search (FOS) and its variants [12, 13, 16], which can form general models using combinations of linear and non-linear basis functions. These methods are based on Gram - Schmidt and Modified Gram - Schmidt orthogonalization as well as a Cholesky type factorization approaches. These methods build the model incrementally based on a re-orthogonalization approach. The time series is projected to this orthogonal set and until a prescribed modelling error is achieved. However, inherently parallel orthogonalization procedures have instabilities and stable variants are inherently sequential, thus cannot take advantage of novel multicore architectures. Furthermore, for models composed of a large number of trigonometric components, retraining the model requires re-computation of the coefficients performed by backward - forwards substitution, which is inherently sequential.

In this article a novel inherently parallel Schur complement based pseudoinverse matrix modelling and forecasting technique is proposed. The proposed scheme is based on a recursive pseudoinverse matrix to form a model based on

a predefined set of linearly independent (linear or non-linear) basis functions. Basis functions are accumulated recursively into the model until a prescribed error is achieved. The stability of the model is ensured by enforcing positive definiteness of the dot product matrix of basis functions through monitoring of error reduction as well as monitoring the magnitude of the diagonal elements. Retraining of the produced model is limited to the pseudoinverse matrix by vector product. By exploiting orthogonality features, computing the residual time series is avoided, improving computational complexity. Discussions and implementation details regarding the computation of the pseudoinverse matrix and model formation are also provided. The case of sinusoidal basis functions is also discussed along with a technique to accurately determine frequencies based on the proposed approach. The efficiency and applicability of the proposed scheme is assessed in a variety of time series with different characteristics.

In Section 2, the recursive Schur complement based pseudoinverse matrix of basis functions is introduced. In Section 3, the Schur complement based pseudoinverse matrix is utilized in the design of the proposed modelling technique in order to computed the weights of the respective basis. Discussions on the stability of the scheme and implementation details are also provided. In Section 4, the case of sinusoidal basis functions is examined and discussed. In Section 5, numerical results are presented depicting the applicability and accuracy of the proposed scheme along with discussions for several time series.

2 Recursive Schur complement based pseudoinverse matrix of basis functions

Let us consider a matrix composed of a set of n basis functions:

$$X = \begin{bmatrix} x_1(t) \ x_2(t) \ \ldots \ x_n(t) \end{bmatrix}, \tag{1}$$

where $x_i(t), 1 \leq i \leq n$ are the basis functions and t is the time variable. We can expand a general time series y based on the $x_i(t)$:

$$y = a_1 x_1(t) + a_2 x_2(t) + \ldots + a_n x_n(t) + \epsilon(t),. \tag{2}$$

where $\epsilon(t)$ is the error term and $a_i, 1 \leq i \leq n$ are the unknown coefficients that have to be determined. The eq. (2) can be written in block form:

$$y = Xa. \tag{3}$$

or equivalently,

$$a = X^+ y = (X^T X)^{-1} X^T y. \tag{4}$$

Thus, the coefficients can be determined by solving the least squares linear system (3) or through the pseudoinverse X^+ (4). However, in many cases the basis functions are not known *a priori* and are computed iteratively or recurrently up to prescribed tolerance. This requires formation and solution of multiple least squares linear systems, which increases substantially the computational work. To

avoid such issues modelling techniques such as Fast Orthogonal Search (FOS) and variants [12, 13, 16] have been proposed, based on Gram - Schmidt and Modified Gram - Schmidt orthogonalization as well as Cholesky factorization type approaches. These methods build the model incrementally by adding basis functions and re-orthogonalize with the already computed ones in order to project the time series and reduce the modelling error.

In order to avoid computational costs and instabilities involved in using orthogonalization procedures, while retaining the flexibility of incrementally building the model, a novel approximate inverse scheme is proposed. Let as consider the matrix X_i containing up to the i-th base and an additional base F:

$$X_{i+1} = \begin{bmatrix} X_i \mid F \end{bmatrix} \tag{5}$$

or:

$$K_{i+1} = X_{i+1}^T X_{i+1} = \begin{bmatrix} X_i^T X_i & X_i^T F \\ F^T X_i & F^T F \end{bmatrix} = \begin{bmatrix} K_i & b \\ b^T & d \end{bmatrix}, \tag{6}$$

where $b = X_i^T F$ and $d = F^T F$. It should be noted that the matrix $X_{i+1}^T X_{i+1}$ is Symmetric Positive Definite and thus invertible, under the assumption that the basis function are linearly independent.

Computing the inverse of the matrix K_{i+1} enables formation of the pseudoinverse X_{i+1}^+ and consequently the computation of the coefficients a_i to form the model. The matrix K_{i+1} can be factored to enable easier inversion and avoid the update step for the elements of the inverse of the matrix K_i, which is essential in case of non-factored inverse matrices [7]:

$$K_{i+1} = \begin{bmatrix} G_i^{-T} & 0 \\ b^T K_i^{-1} & 1 \end{bmatrix} \begin{bmatrix} D_i & 0 \\ 0 & s_i \end{bmatrix} \begin{bmatrix} G_i^{-1} & K_i^{-1} b \\ 0 & 1 \end{bmatrix} \tag{7}$$

or equivalently:

$$K_{i+1}^{-1} = \begin{bmatrix} G_i & -K_i^{-1} b \\ 0 & 1 \end{bmatrix} \begin{bmatrix} D_i^{-1} & 0 \\ 0 & s_i^{-1} \end{bmatrix} \begin{bmatrix} G_i^T & 0 \\ -b^T K_i^{-1} & 1 \end{bmatrix} = G_{i+1} D_{i+1}^{-1} G_{i+1}^T \tag{8}$$

where the matrix D_i retains the Schur complements s_i corresponding to each addition of a basis functions. The Schur complements are of the form: $s_i = d - b^T G_i D_i^{-1} G_i^T b$. Due to the symmetric nature of the the inverse K_{i+1}^{-1} only the factors D_{i+1} and G_{i+1} are required. Thus, we have:

$$G_{i+1} = \begin{bmatrix} G_i & -G_i D_i^{-1} z \\ 0 & 1 \end{bmatrix} \;\; and \;\; D_{i+1}^{-1} = \begin{bmatrix} D_i^{-1} & 0 \\ 0 & (d - z^T D_i^{-1} z)^{-1} \end{bmatrix} \tag{9}$$

where $z = G_i^T b$. Addition of a new basis is limited to simple "matrix times vector" operations, which can be efficiently performed in modern CPUs or accelerators such as GPUs. The storage requirements for the inverse are limited to the upper triangular matrix G_i, which retains $i(i+1)/2$ elements, and the diagonal matrix D_i retaining i elements.

In order to avoid breakdowns in case of weak linear independence between different basis functions or loss of numerical precision, a condition on the magnitude of the diagonal element (Schur Complement) should be imposed. If the diagonal element is close to machine precision, in practice less than $\sqrt{\epsilon_{mach}}$, it is substituted with $\sqrt{\epsilon_{mach}}$, thus avoiding breakdowns.

3   Schur based pseudoinverse matrix modelling and forecasting

The choice of basis functions for creating a model can be arbitrary, since they are evaluated and stored in the columns of matrix X. However, the amount and type of functions chosen affects the accuracy of the model and computational work. Considering the basis functions of eq. (1), the general time - series of eq. (2) and the recursive inverse matrix of eq. (8) we can obtain the coefficients a_i, after the addition of the $i + 1$ basis function, as follows:

$$\begin{bmatrix} a_i \\ b \end{bmatrix} = \begin{bmatrix} G_i & -G_i D_i^{-1} G_i^T X_i^T F \\ 0 & 1 \end{bmatrix} \begin{bmatrix} D_i^{-1} & 0 \\ 0 & s_i^{-1} \end{bmatrix} \begin{bmatrix} G_i^T & 0 \\ -F^T X_i G_i D_i^{-1} G_i^T & 1 \end{bmatrix} \begin{bmatrix} X_i^T y \\ F^T y \end{bmatrix} \quad (10)$$

where $s_i = F^T F - F^T X_i G_i D_i^{-1} G_i^T X_i^T F$. The addition of a basis function involves updating of the values of already computed coefficients a_i corresponding to the i basis functions. Let $g_{i+1} = -G_i D_i^{-1} G_i^T X_i^T F$ then we have:

$$a_{i+1} = \begin{bmatrix} a_i^* \\ b \end{bmatrix} = \begin{bmatrix} a_i + g_{i+1} b \\ s_i^{-1}(F^T + g_{i+1}^T X_i^T)y \end{bmatrix} \quad (11)$$

where $a_i = G_i D_i^{-1} G_i^T X_i^T y$ and $s_i = F^T(F + X_i g_{i+1})$. The coefficients a_i^* are updated after the inclusion of basis function F, while matrix X_i retains all basis functions up to i-th. The residual time series update equation, with respect to the addition of a new basis function can be formed using eq. (11) as follows:

$$\begin{bmatrix} X_i \ F \end{bmatrix} a_{i+1} = \begin{bmatrix} X_i \ F \end{bmatrix} \begin{bmatrix} a_i^* \\ b \end{bmatrix} \quad (12)$$

or:

$$r_{i+1} = y - \begin{bmatrix} X_i \ F \end{bmatrix} \begin{bmatrix} a_i^* \\ b \end{bmatrix} = y - X_i a_i^* - Fb = r_i^* - Fb = r_i - (X_i g_{i+1} + F)b. \quad (13)$$

Using eq. (13) the norm of the residual, after addition of a basis function, can be computed as follows:

$$r_{i+1} = r_i - (X_i g_{i+1} + F)b = r_i - (I - X_i G_i D_i^{-1} G_i^T X_i^T)Fb = r_i - (I - P_{X_i})Fb, \quad (14)$$

where P_{X_i} is an orthogonal projection operator onto the subspace spanned by the columns of X_i. The eq. (14) implies also the following:

$$P_{X_i} r_i = 0 \ \ or \ \ r_i \perp span(X_i). \quad (15)$$

Thus, we have:

$$\|r_{i+1}\|_2^2 = \|r_i\|_2^2 - \|(X_i g_{i+1} + F)b\|_2^2 = \|r_i\|_2^2 - b^2 s_i, \tag{16}$$

or

$$\|r_{i+1}\|_2 = \sqrt{\|r_i\|_2^2 - b^2 s_i}. \tag{17}$$

Eq. (17) can be used to assess progress of fitting, instead of computing the norm of the residual at each iteration, substantially improving performance especially in the case of increased number of basis functions. Moreover, the quantity $b^2 s$ is positive since matrix $X_i^T X_i$ is Symmetric Positive Definite, thus its Schur Complement is also Symmetric Positive Definite, leading to monotonic reduction of the norm of the residual r_{i+1}.

Algorithm 1 Pseudoinverse Matrix Modelling
$(a = pmm(y, F_i, \epsilon))$

1: **Let** y $(T \times 1)$ be the input time series of length T, X_i is the matrix retaining the up to the i-th basis function, F_{i+1} the $i+1$ basis function, M is the maximum number of basis functions, a_i the coefficients corresponding to the i-th basis function and ϵ is the prescribed tolerance.
2: $G_1 = 1$, $D_1 = (F_1^T F_1)$, $s = F_1^T F_1$
3: $b = s^{-1} F_1^T y$
4: $\rho = \|y\|_2^2$, $\rho_0 = \rho$
5: $a_1 = b$
6: $\rho = \rho - b^2 s$
7: $X_1 = F_1$
8: **for** $i = 1$ to $M - 1$ **do**
9:     $g = -G_i D_i^{-1} G_i^T X_i^T F_{i+1}$
10:     $F^* = F_{i+1} + X_i g$
11:     $s = F_{i+1}^T F^*$
12:     **if** $|s| < \sqrt{\epsilon_{mach}}$ **then**
13:         $s = \sqrt{\epsilon_{mach}}$
14:     **end if**
15:     $b = s^{-1} (F^*)^T y$
16:     $a_{1:i} = a_{1:i} + gb$, $a_{i+1} = b$
17:     $G_{i+1} = \begin{bmatrix} G_i & g \\ 0 & 1 \end{bmatrix}$
18:     $D_{i+1} = s$, $X_{i+1} = F_{i+1}$
19:     $\rho = \rho - b^2 s$
20:     **if** $\sqrt{\rho} < \epsilon \sqrt{\rho_0}$ **then**
21:         **return** $a_{1:i+1}$
22:     **end if**
23: **end for**

The initial conditions for eqs. (11) and (13) are the following:

$$G_1 = 1 \tag{18}$$

$$D_1^{-1} = (X_1^T X_1)^{-1} \tag{19}$$

and

$$a_1 = (X_1^T X_1)^{-1} X_1^T y. \tag{20}$$

The procedure for computing the coefficients a_i is described in the Alg. 1. The chosen basis functions are evaluated for $t = 1, ..., T$ and stored in the columns of matrix X. Thus, the model can be formed by computing Xa. The formed model can be used to compute forecasts up to predefined horizon h. This can be achieved by initially progressing the set of basis function in time, e.g. evaluate them for $t^* = T + 1, ..., T + h$ and store them to the columns of a matrix $X^* = [x_1(t^*) \ldots x_n(t^*)]$. Then, the forecasts can be computed by $X^* a$. In practice, the choice of basis functions can be arbitrary, e.g. linear or non-linear or combinations.

The stability of the model with respect to each addition of a basis function can be ensured by allowing basis function to be added only if the error reduction, caused by such an addition, is positive ($b^2 s \geq 0$) and the reduction of the error does not render the error term negative $\rho < 0$ ($\rho \geq b^2 s$). These conditions ensure invertibility of the matrix $X_i^T X_i$ and positive definiteness of the inverse matrix $G_i D_i^{-1} G_i^T$, by allowing inclusion of basis functions that are suitably linearly independent to the already selected ones.

4   The case of sinusoidal basis functions

Sinusoidal basis functions can be used to form a model for general time series. In the case of strong trigonometrical phenomena in a time series, such basis functions can be used to capture them. The sinusoidal basis functions are of the following form:

$$b_i(t) = A cos(\omega_i t) + B sin(\omega_i t), \tag{21}$$

where ω_i is the frequency. Estimation of frequencies can be performed using techniques such as the Fast Fourier Transform through the spectrum or the Quinn - Fernandes algorithm [22]. The proposed scheme allows for determining frequencies with arbitrary accuracy through frequency searching similar to the procedure described in [12]. In case of the proposed technique basis function of the form $F = cos(\omega_i t)$ followed by a basis function of the form $F = sin(\omega_i t)$ for various frequencies $\omega_i \in (0, pi]$ are fitted and the frequency that results in maximum error reduction is selected. The selected frequency becomes part of the model, is excluded from the search space, and the procedure continues until the error criterion is met. The search space $(0, pi)$ is sampled based on a prescribed sampling interval $\delta\omega$. The choice of $\delta\omega$ affects the accuracy in which the frequencies are determined. The advantages of this technique are that frequencies can be determined in parallel while the residual time series need not be computed explicitly.

In order to assess the accuracy of the technique the following example is provided. Let us consider the following:

$$y(t) = cos(\omega_1 t) + 3sin(\omega_2 t) + 2cos(\omega_3 t) - cos(\omega_4 t) + sin(\omega_4 t) + \sigma(t) \tag{22}$$

where $\sigma(t)$ is following uniform random distribution with average equal to 0.5. The frequencies ω_i are: $\omega_1 = 0.0546$, $\omega_2 = 0.83120$, $\omega_3 = 1.87120$ and $\omega_4 = 1.91320$. The frequencies are estimated using Fast Fourier Transform (FFT), Quinn - Fernandes method based on FFT as initial guess coupled with Ordinary Least Squares (OLS) method and the proposed technique. The results are given in Table 1.

Table 1: Percentage errors for frequency estimation.

Frequency	FFT	QF-OLS	$\delta\omega = 10^{-4}$	$\delta\omega = 10^{-5}$
ω_1	3.5690	0.0250	0.1167	0.0592
ω_2	0.2189	0.0007	0.0079	0.0003
ω_3	0.0636	0.0091	0.0132	0.0048
ω_4	0.1626	0.0021	0.0016	0.0017

The choice of $\delta\omega$, in the proposed scheme, should be less than the sampling interval of the Fast Fourier Transform, e.g. $\delta\omega \leq \frac{2*pi}{N}$, in order to allow accurate determination of the frequencies and avoid undersampling.

The proposed technique can be used to estimate the frequency to improved accuracy compared to FFT or the QF-OLS method. The proposed scheme can be coupled also with either the FFT or QF-OLS method and hybrid schemes can be designed leveraging the advantages of those schemes. This will be studied in future work.

5   Numerical results

In this section the applicability and accuracy of the proposed scheme is examined by applying the proposed technique to three time series. The two error measures used to assess the forecasting error was Mean Absolute Percentage Error (MAPE) and Mean Absolute Deviation (MAD):

$$MAPE = \frac{100}{T} \sum_{i=1}^{T} \frac{|y_i - \hat{y}_i|}{|y_i|} \text{ and } MAD = \frac{1}{T} \sum_{i=1}^{T} |y_i - \hat{y}_i|, \qquad (23)$$

where y_i are the actual values, $\hat{y}_i$ the forecasted values and T the length of the test set. The basis functions chosen to model the selected time series were:

$$F_1 = 1, F_2 = t, F_3 = e^t, F_i = A\cos(\omega_i t) + B\sin(\omega_i t), i \geq 4, t \geq 0. \qquad (24)$$

The linear and exponential basis were added to automatically capture such trends in the data. It should be mentioned that the time variable t is scaled for the linear and exponential components to improve numerical behavior during inversion of the matrix of basis functions.

5.1   US airline passenger volume

This US airline passenger volume dataset was extracted from R Studio and is composed of monthly total volumes of passengers spanning from January 1949 to December 1960 (144 samples). The training part was composed of 75% of the dataset, while the test part was composed of 25% of the dataset, specifically the training part included 108 samples and the test part included 36 samples, as presented in Fig. 1. The prescribed interval $\delta\omega$ for frequency search was set to 0.001 and the prescribed tolerance for fitting the model was set to $\epsilon = 0.01$. The forecasted values along with the actuals are given in Fig. 2. The MAPE and MAD of the forecasts were 9.3474 and 40.0410, respectively. From Fig. 2 we observe that the proposed scheme captured the exponential and linear tendency automatically as well as the underlying trigonometric phenomena, without requiring any pre-processing steps for the input data apart from maximum scaling.
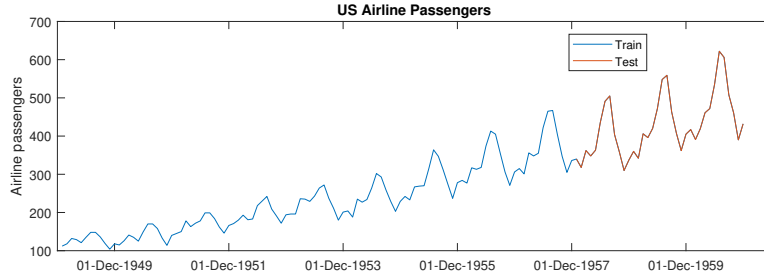
Fig. 1: Train and test parts for the US airline passenger volume dataset.
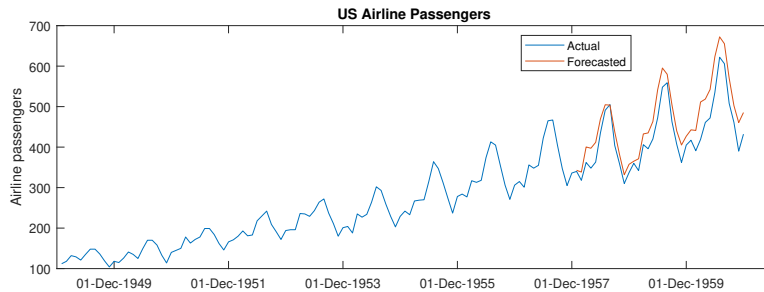

Fig. 2: Forecasted and actual values for the US airline passenger volume dataset.

With respect to the value of the coefficients comprising the model the time series exhibits a significant exponential component, a weak linear component

along with a strong low frequency harmonic component, because of the yearly periodicity of the time series.

5.2   Monthly expenditure on eating out in Australia

The monthly expenditure on eating out in Australia dataset was extracted from R Studio and is composed of the monthly expenditure on cafes, restaurants and takeaway food services in Australia in billion dollars. The dataset is composed of 426 samples spanning a period from April 1982 to September 2017. The training part was composed of $\approx 80\%$ of the dataset while the test part was composed of $\approx 20\%$ of the dataset, specifically the training part included 342 samples and the test part included 84 samples, as presented in Fig. 3. The prescribed interval $\delta\omega$ for frequency search was set to 0.001 and the prescribed tolerance for fitting the model was set to $\epsilon = 0.01$. The forecasted values along with the actuals are given in Fig. 4. The MAPE and MAD of the forecasts were 4.5292 and 0.1465, respectively.
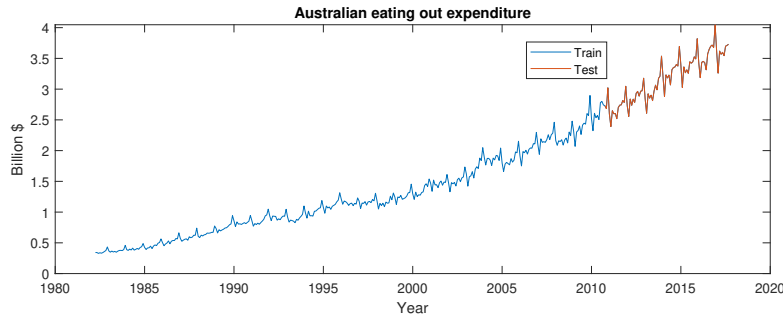

Fig. 3: Train and test parts for the monthly expenditure on eating out in Australia dataset.
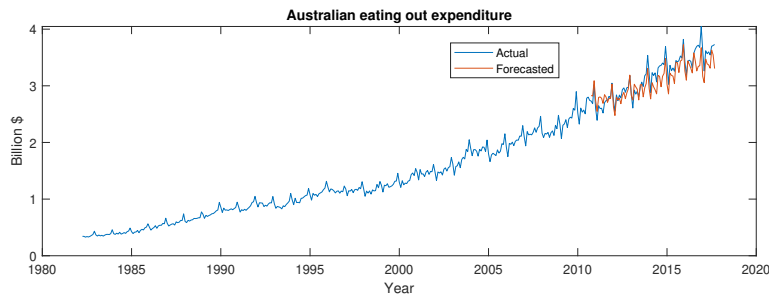

Fig. 4: Forecasted and actual values for the US airline passenger volume dataset.

With respect to the value of the coefficients comprising the model the time series exhibits a strong exponential component, a strong linear component along with a relatively significant medium frequency harmonic components.

5.3 Call volume for a large North American bank

The call volume for a large North American bank dataset was extracted from R Studio and is composed of the volume of calls, per five minute intervals, spanning 164 days starting from 3 March 2003. The dataset is composed of 27716 samples. The training part was composed of $\approx 80\%$ of the dataset while the test part was composed of $\approx 20\%$ of the dataset, specifically the training part included 22325 samples and the test part included 5391 samples, as presented in Fig. 5. The prescribed interval $\delta\omega$ for frequency search was set to 10^{-5} and the prescribed tolerance for fitting the model was set to $\epsilon = 0.088$. The value of the tolerance ϵ is chosen as below that margin the rate of error reduction slows down significantly due to the presence of noise in the form of a large number of frequencies with the same magnitude in the spectrum. This issue can be overcome by increasing the samples of the spectrum, however this substantially increases the computational work, without significant improvement in the forecasting error. The forecasted values along with the actuals are given in Fig. 6. The MAPE and MAD of for the forecasts were 15.1580 and 25.4576, respectively.
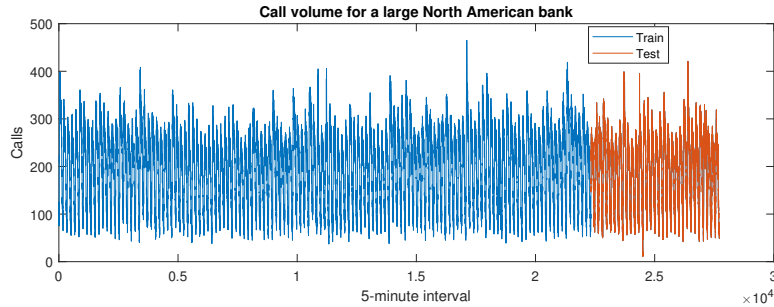

Fig. 5: Train and test parts for the call volume for a large North American bank dataset.

With respect to the value of the coefficients the model has a weak exponential component that is counteracted by a weak linear component. There are also strong low frequency components that contribute significantly in the reduction of the error.

5.4 Discussions

The proposed scheme was able to capture the dominant characteristics of the different time series. The choice of the basis functions substantially affects the
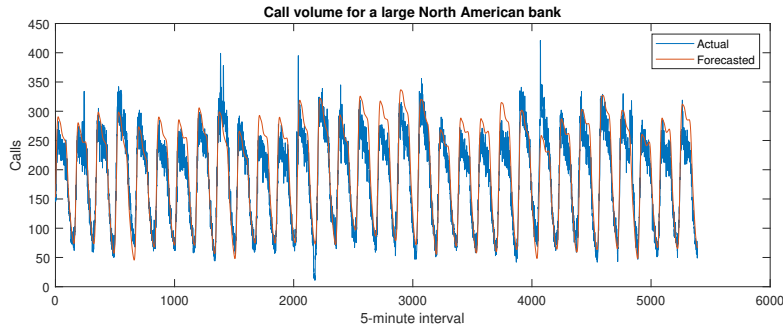
Fig. 6: Forecasted and actual values for call volume for a large North American bank dataset.

estimation and the forecasting error. For example, for the model problem of Subsection 5.3 the linear and exponential basis do not contribute significantly to the accuracy of the model, while also increasing the computational complexity since the dimensions of the pseudoinverse matrix grow. However, to preserve generality and wide applicability, a common set of basis functions was retained for all experiments.

Another important issue is the estimation of frequencies, which for the low value of the $\delta\omega$ parameter requires substantial computational work especially in the case of large training data. In order to reduce computational complexity, frequency estimation can be performed by means of either FFT or the Quinn - Fernandes algorithm [22] or hybrid approaches which will be studied in future research.

The generality of the approach allows the incorporation of basis functions based on nonlinear modelling techniques such as Artificial Neural Networks (ANN) and Support Vector Machines (SVM) trained by subsets of the available dataset. The effect of such basis functions will be studied also in future research.

6   Conclusion

A novel Schur complement based pseudoinverse matrix approach for modelling and forecasting general time series has been proposed. The proposed technique can incorporate linear and non-linear components during model formation, thus avoiding preprocessing and transformation of the time series or restrictive assumptions related to the statistical properties of the data. Stability of the model is ensured by enforcing positive definiteness of the dot product matrix of basis functions $(X^T X)$ and its inverse, and monotonic reduction of the error. A frequency detection technique is also presented based on the proposed scheme. The proposed scheme does not require preprosessing of time series and is assessed by modelling several time series exhibiting combinations of exponential, linear,

trigonometric and random characteristics. Moreover, the model relies on a single parameter and it is suitable for modelling general time series.

Future work is directed towards the design of a parallel approximate pseudoinverse matrix approach in order to reduce storage requirements especially in the case of large number of basis functions. Moreover, an adaptive approach for frequency estimation is under further research.

Acknowledgement

References

1. Awartani, B.M., Corradi, V.: Predicting the volatility of the S&P-500 stock index via GARCH models: the role of asymmetries. International Journal of Forecasting **21**(1), 167–183 (2005)
2. Box, G.E.P., Jenkins, G.M.: Time Series Analysis Forecasting and Control. San Francisco: Holden Day (1976)
3. Brown, R.G.: Smoothing, Forecasting and prediction of discrete time series. Englewood Cliffs, NJ, Prentice Hall (1963)
4. Cleveland, R.B., Cleveland, W.S., McRae, J.E., Terpenning, I.: Stl: A seasonal-trend decomposition procedure based on loess (with discussion). Journal of Official Statistics **6**, 3–73 (1990)
5. Dagum, E.B.: Revisions of time varying seasonal filters. Journal of Forecasting **1**(2), 173–187 (1982). https://doi.org/10.1002/for.3980010204
6. Engle, R.F.: Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica **50**(4), 987–1007 (1982)
7. Filelis-Papadopoulos, C.K.: Incomplete inverse matrices. Numerical Linear Algebra with Applications p. Accepted (2021). https://doi.org/https://doi.org/10.1002/nla.2380
8. Gooijer, J.G., Hyndman, R.: 25 years of iif time series forecasting: A selective review. Monash Econometrics and Business Statistics Working Papers 12/05, Monash University, Department of Econometrics and Business Statistics (2005), https://EconPapers.repec.org/RePEc:msh:ebswps:2005-12
9. Harrison, P.J., Stevens, C.F.: Bayesian forecasting. Journal of the Royal Statistical Society. Series B (Methodological) **38**(3), 205–247 (1976)
10. Harvey, A.C.: Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge: Cambridge University Press (1989)
11. Holt, C.C.: Forecasting seasonals and trends by exponentially weighted moving averages. International Journal of Forecasting **20**, 5–13 (2004)
12. Korenberg, M.J., Paarmann, L.D.: Orthogonal approaches to time-series analysis and system identification. IEEE Signal Processing Magazine **8**(3), 29–43 (1991)
13. Li, X., Tian, J., Wang, X., Dai, J., Ai, L.: Fast orthogonal search method for modeling nonlinear hemodynamic response in fMRI. In: Amini, A.A., Manduca, A. (eds.) Medical Imaging 2004: Physiology, Function, and Structure from Medical Images. vol. 5369, pp. 219 – 226. International Society for Optics and Photonics, SPIE (2004). https://doi.org/10.1117/12.536165, https://doi.org/10.1117/12.536165

14. Makridakis, S., Hibon, M.: Exponential smoothing: The effect of initial values and loss function on post-sample forecasting accuracy. International Journal of Forecasting **7**, 317–330 (1991)

15. Makridakis, S., Spiliotis, E., Assimakopoulos, V.: The m4 competition: 100,000 time series and 61 forecasting methods. International Journal of Forecasting **36**(1), 54 – 74 (2020). https://doi.org/https://doi.org/10.1016/j.ijforecast.2019.04.014, m4 Competition

16. Osman, A., Afan, H.A., Allawi, M.F., Jaafar, O., Noureldin, A., Hamzah, F.M., Ahmed, A.N., El-shafie, A.: Adaptive fast orthogonal search (fos) algorithm for forecasting streamflow. Journal of Hydrology **586**, 124896 (2020). https://doi.org/https://doi.org/10.1016/j.jhydrol.2020.124896, http://www.sciencedirect.com/science/article/pii/S0022169420303565

17. Pagan, A.: The econometrics of financial markets. Journal of Empirical Finance **3**(1), 15–102 (1996)

18. Parzen, E.: Ararma models for time series analysis and forecasting. Journal of Forecasting **1**(1), 67–82 (1982). https://doi.org/10.1002/for.3980010108, https://onlinelibrary.wiley.com/doi/abs/10.1002/for.3980010108

19. Poskitt, D., Tremayne, A.: The selection and use of linear and bilinear time series models. International Journal of Forecasting **2**(1), 101 – 114 (1986). https://doi.org/https://doi.org/10.1016/0169-2070(86)90033-6

20. Proietti, T.: Comparing seasonal components for structural time series models. International Journal of Forecasting **16**(2), 247 – 260 (2000). https://doi.org/https://doi.org/10.1016/S0169-2070(00)00037-6

21. Quenouille, M.H.: The Analysis of Multiple Time-Series. London: Griffin, 2nd edn. (1968)

22. Quinn, B.G., Fernandes, J.M.: A fast efficient technique for the estimation of frequency. Biometrika **78**(3), 489–497 (1991)

23. Riise, T., Tjozstheim, D.: Theory and practice of multivariate arma forecasting. Journal of Forecasting **3**(3), 309–317 (1984). https://doi.org/10.1002/for.3980030308, https://onlinelibrary.wiley.com/doi/abs/10.1002/for.3980030308

24. Tashman, L.J.: Out-of-sample tests of forecasting accuracy: an analysis and review. International Journal of Forecasting **16**(4), 437–450 (2000)

25. Taylor, J.W.: Exponential smoothing with a damped multiplicative trend. International Journal of Forecasting **19**, 273–289 (2003)

26. Wiener, N.: Non-linear Problems in Random Theory. London: Wiley (1958)

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