18^{th} *World IMAS/MODSIM Congress, Cairns, Australia 13-17 July 2009*
http://mssanz.org.au/modsim09

Nonparametric time series forecasting with dynamic updating

Han Lin Shang and Rob. J. Hyndman

Department of Econometrics & Business Statistics, Monash University, Clayton, Victoria
Email: HanLin.Shang@buseco.monash.edu.au

Abstract: We present a nonparametric method to forecast a seasonal time series, and propose four dynamic updating methods to improve point forecast accuracy. Our forecasting and dynamic updating methods are data-driven and computationally fast, and they are thus feasible to be applied in practice. We will demonstrate the effectiveness of these methods using monthly El Niño time series from 1950 to 2008 (http://www.cpc.noaa.gov/data/indices/sstoi.indices).

Let $\{Z_w, w \in [0, \infty)\}$ be a seasonal univariate time series which has been observed at N equispaced time. Aneiros-Pérez & Vieu (2008) assume that N can be written as $N = np$, where n is the number of samples and p is dimensionality. To clarify this, in the El Niño time series from 1950 to 2008, we have $N = 708, n = 59, p = 12$. The observed time series $\{Z_1, \cdots, Z_{708}\}$ can thus be divided into 59 successive paths of length 12 in the following setting: $\boldsymbol{y}_t = \{Z_w, w \in (p(t-1), pt]\}$, for $t = 1, \cdots, 59$. The problem is to forecast future processes, denoted as $\boldsymbol{y}_{n+h,h>0}$, from the observed data.

To solve this problem, we apply a nonparametric method known as principal component analysis (PCA) to decompose a complete (12×59) data matrix ($\boldsymbol{Y} = [\boldsymbol{y}_1, \cdots, \boldsymbol{y}_{59}]$) into a number of principal components and their associated principal component scores. That is,

$$\boldsymbol{Y} = \hat{\boldsymbol{\mu}} + \hat{\boldsymbol{\phi}}_1'\hat{\boldsymbol{\beta}}_1 + \cdots + \hat{\boldsymbol{\phi}}_K'\hat{\boldsymbol{\beta}}_K + \hat{\boldsymbol{\epsilon}}$$

where $\hat{\boldsymbol{\mu}} = [\hat{\mu}_1, \cdots, \hat{\mu}_{12}]'$ is the pointwise mean vector; $\hat{\boldsymbol{\phi}}_1, \cdots, \hat{\boldsymbol{\phi}}_K \in R^K$ ($\hat{\boldsymbol{\phi}}_k = [\phi_{1,k}, \cdots, \phi_{12,k}]$) are estimated principal components; $\hat{\boldsymbol{\beta}}_1, \cdots, \hat{\boldsymbol{\beta}}_K$ ($\hat{\boldsymbol{\beta}}_k = [\beta_{1,k}, \cdots, \beta_{59,k}]'$) are uncorrelated principal component scores satisfying $\sum_{k=1}^K \hat{\boldsymbol{\beta}}_k^2 < \infty$, for $k = 1, \cdots, K$; $\hat{\boldsymbol{\epsilon}}$ is assumed to be a zero-mean 12×59 residual matrix; and $K < 12$ is the optimal number of components.

Since $\hat{\boldsymbol{\beta}}_1, \cdots, \hat{\boldsymbol{\beta}}_K$ are uncorrelated, we can forecast them using a univariate time series (TS) method, like exponential smoothing (Hyndman et al., 2008). Conditioning on the observed data ($\boldsymbol{\mathcal{I}}$) and fixed principal components ($\boldsymbol{\Phi} = \boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_K$), and the forecasted curves are given as

$$\hat{\boldsymbol{y}}_{n+h|n} = \mathrm{E}(\boldsymbol{y}_{n+h}|\boldsymbol{\mathcal{I}}, \boldsymbol{\Phi}) = \hat{\boldsymbol{\mu}} + \hat{\boldsymbol{\phi}}_1'\hat{\beta}_{1,n+h|n} + \cdots + \hat{\boldsymbol{\phi}}_K'\hat{\beta}_{K,n+h|n}, \tag{1}$$

where $\hat{\beta}_{k,n+h|n}$, $k = 1, \cdots, K$ are the forecasted principal component scores.

An interesting problem arises when $N \neq np$, which is an assumption made in Aneiros-Pérez & Vieu (2008). In other words, there are partially observed data in the final year. This motivates us to develop four dynamic updating methods, not only to update our point forecasts, but also to eliminate the assumption in Aneiros-Pérez & Vieu (2008).

Four dynamic updating methods are called the block moving (BM), ordinary least squares (OLS), penalized least squares (PLS), and ridge regression (RR). The BM approach rearranges the observed data matrix to form a complete data matrix by sacrificing some observations in the first year, thus (1) can still be applied. The OLS method considers the partially observed data in the final year as responses, and use them to regress against the corresponding principal components, but it fails to consider historical data. The PLS method effectively combines the advantages of both TS and OLS methods, while the RR method is a well-known shrinkage method for solving ill-posed problems.

Keywords: El Niño time series, penalized least squares, principal component regression.

1 Introduction

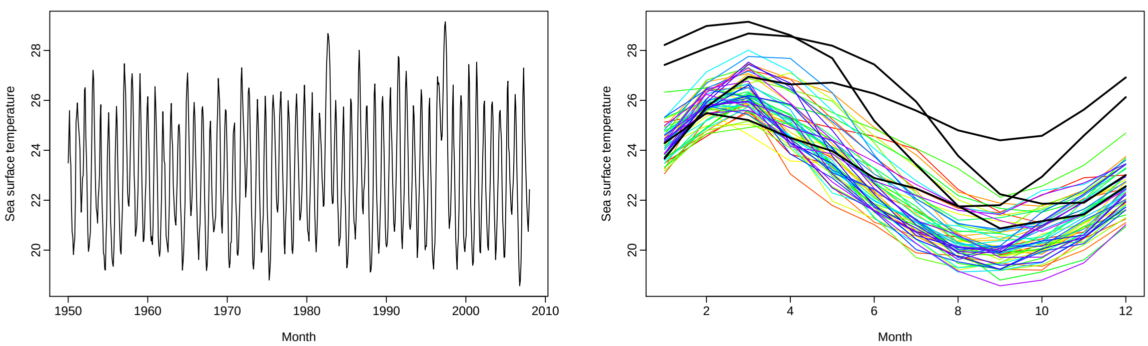
Predicting time series of future values is of prime interest in statistics. Regardless of the kind of statistical modeling used, an important parameter that has to be chosen is the number of past values to use to construct a prediction method. The fewer the number of past predictors, the less flexible the model will be. This well-known phenomenon is particularly troublesome in nonparametric statistics for which the asymptotic behavior of the estimates is exponentially decaying with the number of past values, which are incorporated in the model.

One way to overcome the problem of incorporating a large number of past values into the statistical model is to use functional ideas. The idea is to divide the observed seasonal time series into a sample of trajectories, and to construct a single past (continuous) trajectory rather than many single past (discrete) values. We are interested in predicting a single continuous curve rather than 12 discrete data points in a year.

Recent development in functional time series forecasting include the functional autoregressive of order 1 (Bosq 2000), and functional kernel regression (Aneiros-Pérez & Vieu 2008), and functional principal component regression (Hyndman & Ullah 2007, Hyndman & Booth 2008). However, to our knowledge, there is little has been done to address the dynamic updating problem when the final curve is partially observed. The contribution of this paper is to propose four dynamic updating methods in a multivariate setting, although the methods can easily be extended to a functional framework using a nonparametric smoothing technique.

This paper is organized as follows: data set is described in Section 2, Section 3 portrays the forecasting methods, while Section 4 introduces our four dynamic updating methods. In Section 5, we compare the point forecast accuracy with several existing methods. Conclusions are presented in Section 6.

2 Data set



(a) A univariate time series display.

(b) A functional time series display.

Figure 1: Exploratory plots of the El Niño data set from Jan 1950 to Dec 2008 measured by moored buoys in the region defined by the coordinate $0 - 10°$ South and $90 - 80°$ West.

We consider the monthly El Niño indices from Jan, 1950 to Dec, 2008, available online at http://www.cpc.noaa.gov/data/indices/sstoi.indices. These El Niño indices are measured by moored buoys in the "Niño region" defined by the coordinates $0 - 10°$ South and $90 - 80°$ West. While a univariate time series display is given in Figure 1a, the monthly data graphed for each year are shown in Figure 1b.

3 Forecasting method

Let $\{Z_w, w \in [0, \infty)\}$ be a seasonal univariate time series which has been observed at N equispaced time. In the El Niño time series from 1950 to 2008, we have $N = 708, n = 59, p = 12$. The observed time series is then divided into 59 successive paths of length 12 in the following setting

$$\boldsymbol{y}_t = \{Z_w, w \in (p(t-1), pt]\} \quad \forall t = 1, \cdots, 59.$$

Our method begins with decentralizing data by subtracting pointwise mean $\hat{\boldsymbol{\mu}} = \frac{1}{p}\sum_{i=1}^{p} \boldsymbol{y}_i$. The mean-adjusted data are denoted as $\hat{\boldsymbol{Y}}^* = \boldsymbol{Y} - \hat{\boldsymbol{\mu}}$, where $\boldsymbol{Y} = [y_1, \cdots, y_{59}]$. Using principal component analysis (PCA), $\hat{\boldsymbol{Y}}^*$ can be approximated by the sum of separable principal components and their associated scores in order to achieve minimal L_2 loss of information. Computationally, applying singular value decomposition to $\hat{\boldsymbol{Y}}^*$ gives $\hat{\boldsymbol{Y}}^* = \boldsymbol{\Psi}\boldsymbol{\Lambda}\boldsymbol{V}^{'}$, where $\boldsymbol{\Psi}$ is an $p \times p$ unitary matrix, $\boldsymbol{\Lambda}$ is a $p \times p$ diagonal matrix, and $\boldsymbol{V}^{'}$ is $p \times n$ matrix. Let $\phi_{k,i}$ be the $(i,k)^{\text{th}}$ element of $\boldsymbol{\Psi}$ truncated at first K columns. Since $\hat{\boldsymbol{B}} = \hat{\boldsymbol{Y}}^{*'}\boldsymbol{\Psi}$, $\hat{\beta}_{t,k}$ is the $(t,k)^{\text{th}}$ element of $\hat{\boldsymbol{B}}$ truncated at first K columns.

Therefore, $\boldsymbol{Y}$ can be written as

$$\boldsymbol{Y} = \hat{\boldsymbol{\mu}} + \hat{\boldsymbol{\phi}}_1'\hat{\boldsymbol{\beta}}_1 + \cdots + \hat{\boldsymbol{\phi}}_K'\hat{\boldsymbol{\beta}}_K + \hat{\boldsymbol{\varepsilon}},$$

where $\hat{\boldsymbol{\mu}} = [\hat{\mu}_1, \cdots, \hat{\mu}_{12}]'$ is the pointwise mean vector; $\hat{\boldsymbol{\phi}}_1, \cdots, \hat{\boldsymbol{\phi}}_K \in R^K$ ($\hat{\boldsymbol{\phi}}_k = [\phi_{1,k}, \cdots, \phi_{12,k}]$) are estimated principal components; $\hat{\boldsymbol{\beta}}_1, \cdots, \hat{\boldsymbol{\beta}}_K$ ($\hat{\boldsymbol{\beta}}_k = [\beta_{1,k}, \cdots, \beta_{59,k}]'$) are uncorrelated principal component scores satisfying $\sum_{k=1}^{K} \hat{\boldsymbol{\beta}}_k^2 < \infty$, for $k = 1, \cdots, K$; $\hat{\boldsymbol{\epsilon}}$ is assumed to be a zero-mean 12×59 residual matrix; and $K < 12$ is the optimal number of components.

Since $\{\hat{\boldsymbol{\beta}}_1, \cdots, \hat{\boldsymbol{\beta}}_K\}$ are uncorrelated to each other, it is appropriate to forecast each series $\hat{\boldsymbol{\beta}}_k$ using a univariate time series forecasting method, such as the ARIMA (Box et al. 2008) or exponential smoothing (Hyndman et al. 2008). By conditioning on the historical observations ($\boldsymbol{\mathcal{I}}$) and the fixed principal components ($\boldsymbol{\Phi} = \hat{\boldsymbol{\phi}}_1, \cdots, \hat{\boldsymbol{\phi}}_K$), the forecasted curves are expressed as

$$\hat{\boldsymbol{y}}_{n+h|n}^{\text{TS}} = \text{E}[\boldsymbol{y}_{n+h}|\boldsymbol{\mathcal{I}}, \boldsymbol{\Phi}] = \hat{\boldsymbol{\mu}} + \hat{\boldsymbol{\phi}}_1'\hat{\beta}_{1,n+h|n} + \cdots + \hat{\boldsymbol{\phi}}_K'\hat{\beta}_{K,n+h|n}, \tag{2}$$

where $\hat{\beta}_{k,n+h|n}$ denotes the h-step-ahead forecasts of $\beta_{k,n+h}$.

4 Updating methods

As we observe some recent data consisting of the first m_0 time period of $\boldsymbol{y}_{n+1}$, denoted by $\boldsymbol{y}_{n+1,e} = [y_{n+1,1}, \cdots, y_{n+1,m_0}]^{'}$, we are interested in forecasting the data in the remaining time period of year $n+1$, as denoted by $\boldsymbol{y}_{n+1,l} = [y_{n+1,m_0+1}, \cdots, y_{n+1,12}]^{'}$. Using (2), the time series (TS) forecast of $\boldsymbol{y}_{n+1,l}$ is given by

$$\hat{\boldsymbol{y}}_{n+1|n,l}^{\text{TS}} = \text{E}[\boldsymbol{y}_{n+1,l}|\boldsymbol{\mathcal{I}}_l, \boldsymbol{\Phi}_l] = \hat{\boldsymbol{\mu}}_l + \hat{\boldsymbol{\phi}}_{1,l}'\hat{\beta}_{1,n+1|n}^{\text{TS}} + \cdots + \hat{\boldsymbol{\phi}}_{K,l}'\hat{\beta}_{K,n+1|n}^{\text{TS}},$$

where $\hat{\boldsymbol{\mu}}_l = [\hat{\mu}_{m_0+1}, \cdots, \hat{\mu}_{12}]'$, $\boldsymbol{\Phi}_l = \{\boldsymbol{\phi}_{1,l}^{'}, \cdots, \boldsymbol{\phi}_{K,l}^{'}\}$ are the principal components corresponding to the remaining time period, $\hat{\beta}_{k,n+1|n}^{\text{TS}}$ are the one-step-ahead forecasted principal component scores, and $\boldsymbol{\mathcal{I}}_l$ denotes the historical data corresponding to the remaining time period.

It is clear that this TS method does not consider any new observations. With the aim to improve forecast accuracy, it is desirable to dynamically update the forecasts for the remaining time period of the year $n+1$ by incorporating the new observations.

4.1 Block moving method

The block moving (BM) method considers the most recent data as the last observation in a complete data matrix. Because time is a continuous variable, we can observe a complete data matrix at any given time interval. Thus, the TS method can still be applied by sacrificing a number of data points in the first year. This loss of data will not affect the parameter estimation as long as the number of curves is large. A concept diagram is presented in Figure 2 to examplify this idea.

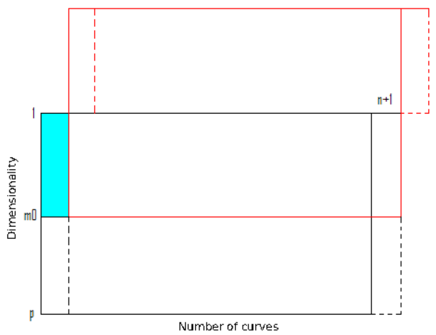


Figure 2: Dynamic update via the block moving approach. The cyan region shows the data loss in the first year. The forecasts for the remaining months in year $n+1$ can be updated by the forecasts using the TS method applied to the upper block.

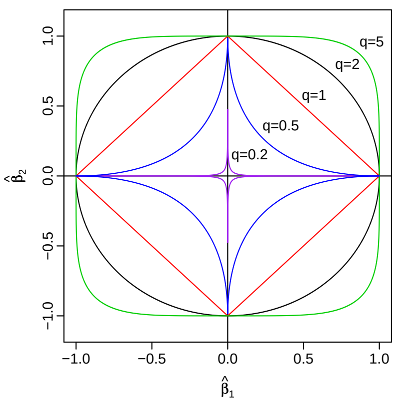


Figure 3: Two-dimensional contours of the symmetric penalty function $p_q(\boldsymbol{\beta}) = |\boldsymbol{\beta}_1|^q + |\boldsymbol{\beta}_2|^q = 1$ for q=0.2, 0.5, 1, 2, 5. The scenario $q = 1$ (red diamond) yields the lasso and $q = 2$ (black circle) yields ridge regression.

4.2 Ordinary least squares approach

We can express (2) in a matrix form. Let $\boldsymbol{F}^e$ be a $m_0 \times K$ matrix whose $(j,k)^{\text{th}}$ entry is $\hat{\phi}_{k,j}$ for $1 \leq j \leq m_0, 1 \leq k \leq K$. Denote $\hat{\boldsymbol{\beta}}_{n+1} = [\hat{\beta}_{1,n+1}, \cdots, \hat{\beta}_{K,n+1}]'$ as a $K \times 1$ vector, and $\hat{\boldsymbol{\epsilon}}_{n+1,e} = [\hat{\epsilon}_{n+1,1}, \cdots, \hat{\epsilon}_{n+1,m_0}]'$ be a $m_0 \times 1$ vector. As the mean-adjusted $\boldsymbol{y}^*_{n+1,e}$ becomes available, we have an ordinary least squares (OLS) regression expressed as

$$\boldsymbol{y}^*_{n+1,e} = \boldsymbol{F}^e \hat{\boldsymbol{\beta}}_{n+1} + \hat{\boldsymbol{\epsilon}}_{n+1,e}.$$

The $\hat{\boldsymbol{\beta}}_{n+1}$ can be approximated via the OLS method by solving

$$\underset{\hat{\boldsymbol{\beta}}_{n+1}}{\operatorname{argmin}} \left\| \hat{\boldsymbol{y}}^*_{n+1,e} - \boldsymbol{F}^e \hat{\boldsymbol{\beta}}_{n+1} \right\|,$$

thus obtaining

$$\hat{\boldsymbol{\beta}}^{\text{OLS}}_{n+1} = \left(\boldsymbol{F}^{e'} \boldsymbol{F}^e \right)^{-1} \boldsymbol{F}^{e'} \hat{\boldsymbol{y}}^*_{n+1,e}.$$

The OLS forecast of $\boldsymbol{y}_{n+1,l}$ is then given by

$$\hat{\boldsymbol{y}}^{\text{OLS}}_{n+1|n,l} = \text{E}[\boldsymbol{y}_{n+1,l}|\boldsymbol{\mathcal{I}}_l, \boldsymbol{\Phi}_l] = \hat{\boldsymbol{\mu}}_l + \hat{\boldsymbol{\phi}}'_{1,l}\hat{\beta}^{\text{OLS}}_{1,n+1} + \cdots + \hat{\boldsymbol{\phi}}'_{K,l}\hat{\beta}^{\text{OLS}}_{K,n+1}.$$

4.3 Penalized least squares approach

The OLS method uses new observations to improve forecast accuracy for the remaining time period of year $n+1$, but it needs a sufficient number of observations (at least equal to K) in order for $\hat{\boldsymbol{\beta}}^{\text{OLS}}_{n+1}$ to be numerically stable. In addition, it does not make use of TS forecasts. To overcome these problems, we adapt the idea of PLS by penalizing the OLS forecasts, which deviate significantly from the TS forecasts. The regression coefficients of PLS are obtained by

$$\underset{\hat{\boldsymbol{\beta}}_{n+1}}{argmin}\, \|\hat{\boldsymbol{y}}^*_{n+1,e} - \boldsymbol{F}^e\hat{\boldsymbol{\beta}}_{n+1}\| + \lambda p_q(\hat{\boldsymbol{\beta}}_{n+1}),$$

with penalty parameter λ and a given penalty function $p_q(\hat{\boldsymbol{\beta}}_{n+1})$ expressed as

$$p_q(\hat{\boldsymbol{\beta}}_{n+1}) = \sum_{j=1}^{m_0} |\hat{\boldsymbol{\beta}}_j - \hat{\boldsymbol{\beta}}^{\text{TS}}_j|^q \leq c,$$

which bounds the L_q norm of parameters in the model (Frank & Friedman 1993). The tuning parameter c controls the size of the hypersphere, and hence, the shrinkage of $\hat{\boldsymbol{\beta}}_j$ toward the $\hat{\boldsymbol{\beta}}^{\text{TS}}_j$. A two-dimensional contours of the penalty function for different values of q is presented in Figure 3.

When $q = 2$, the PLS method corresponds to ridge regression, which is rotationally invariant hypersphere centered at the origin. Thus, the $\hat{\boldsymbol{\beta}}_{n+1}$ obtained from the PLS method minimizes

$$\left(\hat{\boldsymbol{y}}^*_{n+1,e} - \boldsymbol{F}^e\hat{\boldsymbol{\beta}}_{n+1}\right)'\left(\hat{\boldsymbol{y}}^*_{n+1,e} - \boldsymbol{F}^e\hat{\boldsymbol{\beta}}_{n+1}\right) + \lambda\left(\hat{\boldsymbol{\beta}}_{n+1} - \hat{\boldsymbol{\beta}}^{\text{TS}}_{n+1|n}\right)'\left(\hat{\boldsymbol{\beta}}_{n+1} - \hat{\boldsymbol{\beta}}^{\text{TS}}_{n+1|n}\right), \tag{3}$$

The first term in (3) measures the goodness of fit, while the second term penalizes the departure of the OLS coefficients from the TS coefficient forecasts. The $\hat{\boldsymbol{\beta}}_{n+1}$ obtained can thus be seen as a tradeoff between these two terms, subject to a penalty parameter λ.

By taking first derivative with respect to $\hat{\boldsymbol{\beta}}_{n+1}$ in (3), we obtain

$$\hat{\boldsymbol{\beta}}^{\text{PLS}}_{n+1} = \frac{\boldsymbol{F}^{e'}\boldsymbol{F}^e}{\boldsymbol{F}^{e'}\boldsymbol{F}^e + \lambda\boldsymbol{I}}\hat{\boldsymbol{\beta}}^{\text{OLS}}_{n+1} + \frac{\lambda\boldsymbol{I}}{\boldsymbol{F}^{e'}\boldsymbol{F}^e + \lambda\boldsymbol{I}}\hat{\boldsymbol{\beta}}^{\text{TS}}_{n+1|n},$$

where $\boldsymbol{I}$ is identity matrix. When the penalty parameter $\lambda \to 0, \hat{\boldsymbol{\beta}}^{\text{PLS}}_{n+1}$ is simply $\hat{\boldsymbol{\beta}}^{\text{OLS}}_{n+1}$; when $\lambda \to \infty$, then $\hat{\boldsymbol{\beta}}^{\text{PLS}}_{n+1}$ reduces to $\hat{\boldsymbol{\beta}}^{\text{TS}}_{n+1|n}$; when $0 < \lambda < \infty$, $\hat{\boldsymbol{\beta}}^{\text{PLS}}_{n+1}$ is a weighted average between $\hat{\boldsymbol{\beta}}^{\text{TS}}_{n+1|n}$ and $\hat{\boldsymbol{\beta}}^{\text{OLS}}_{n+1}$. Therefore, the PLS forecast of $\boldsymbol{y}_{n+1,l}$ is given as:

$$\hat{\boldsymbol{y}}^{\text{PLS}}_{n+1,l} = \text{E}[\boldsymbol{y}_{n+1,l}|\boldsymbol{\mathcal{I}}_l, \boldsymbol{\Phi}_l] = \hat{\boldsymbol{\mu}}_l + \hat{\boldsymbol{\phi}}'_{1,l}\hat{\beta}^{\text{PLS}}_{1,n+1} + \cdots + \hat{\boldsymbol{\phi}}'_{K,l}\hat{\beta}^{\text{PLS}}_{K,n+1}.$$

Similar to (3), ridge regression (RR) penalizes the OLS coefficients, which deviate significantly from 0. The RR coefficients minimize a penalized residual sum of square,

$$\left(\hat{\boldsymbol{y}}^*_{n+1,e} - \boldsymbol{F}^e\hat{\boldsymbol{\beta}}_{n+1}\right)'\left(\hat{\boldsymbol{y}}^*_{n+1,e} - \boldsymbol{F}^e\hat{\boldsymbol{\beta}}_{n+1}\right) + \lambda\left(\hat{\boldsymbol{\beta}}'_{n+1}\hat{\boldsymbol{\beta}}_{n+1}\right), \tag{4}$$

where $\lambda > 0$ is a tuning parameter that controls the amount of shrinkage. By taking the derivative with respect to $\hat{\boldsymbol{\beta}}_{n+1}$ in (4), we obtain

$$\hat{\boldsymbol{\beta}}^{\text{RR}}_{n+1} = \left(\boldsymbol{F}^{e'}F^e + \lambda\boldsymbol{I}\right)^{-1}F^{e'}\boldsymbol{y}^*_{n+1,e}$$

Therefore, the RR forecast of $\boldsymbol{y}_{n+1,l}$ is given as:

$$\hat{\boldsymbol{y}}^{\text{RR}}_{n+1,l} = \text{E}[\boldsymbol{y}_{n+1,l}|\boldsymbol{\mathcal{I}}_l, \boldsymbol{\Phi}_l] = \hat{\boldsymbol{\mu}}_l + \hat{\boldsymbol{\phi}}'_{1,l}\hat{\beta}^{\text{RR}}_{1,n+1} + \cdots + \hat{\boldsymbol{\phi}}'_{K,l}\hat{\beta}_{K,n+1}.$$

5 Point forecast comparison

By means of comparison, we also investigate the forecast performance of seasonal autoregressive moving average (SARIMA), random walk (RW) and mean predictor (MP) models. The MP model consists in predicting values at year $t+1$ by the empirical mean values from year 1 to t for each month, while the RW approach predicts new values at year $t+1$ by the observed values of El Niño indices at year t. The SARIMA requires the specification orders of seasonal and nonseasonal components of an ARIMA model, which are determined by an automatic algorithm of Hyndman & Khandakar (2008). As a result, the optimal SARIMA is an autoregressive model at lag 2 and differencing at lag 12, and a moving average model at lag 1 and differencing at lag 12.

As the presence of outliers can affect the forecast accuracy, we applied an outlier detection method of Hyndman & Shang (2008), and detected four outliers corresponding 1982, 1983, 1997 and 1998. For details on detected outliers, see Dioses et al. (2002). Consequently, the data in years 1982, 1983, 1997 and 1998 are removed from further analysis.

We split the data into a training set (containing indices from 1950 to 1970 excluding the outliers), a validation set (containing indices from 1971 to 1992) for finding optimal λ and K, and a testing set (containing indices from 1993 to 2008). In this data set, the optimal K is determined to be 5.

The best forecast method is determined with a minimal mean square error (MSE) in the testing set. The MSE is expressed as:

$$\text{MSE} = \frac{1}{s}\frac{1}{p}\sum_{w=1}^{s}\sum_{j=1}^{p}(y_{n+w,j} - \hat{y}_{n+w,j})^2,$$

where s represents the length of hold-out testing sample.

Update month	MP	RW	SARIMA	TS	OLS	BM	PLS	RR
Mar-Dec	0.6928	1.3196	1.4155	0.7101	0.8324	0.6895	**0.6623**	0.8356
Apr-Dec	0.7115	1.3607	1.4706	0.7296	0.7147	0.7180	0.6924	**0.6505**
May-Dec	0.6822	1.3683	1.3195	0.7025	1.2033	0.6903	0.6588	**0.6164**
Jun-Dec	0.6792	1.3710	1.1880	0.7036	1.6853	0.6811	0.6111	**0.5868**
Jul-Dec	0.6984	1.4660	1.2089	0.7322	1.4431	0.6772	0.5492	**0.5101**
Aug-Dec	0.7011	1.5726	1.1279	0.7541	1.5444	0.6835	0.6697	**0.6552**
Sep-Dec	0.7056	1.6499	1.0624	0.7800	1.7000	0.7096	**0.6132**	0.6259
Oct-Dec	0.7261	1.6972	**0.5394**	0.8262	0.6713	0.7443	0.6484	0.5653
Nov-Dec	0.7112	1.5097	0.4244	0.8201	0.1151	0.7566	0.4186	**0.1085**
Dec	0.5646	1.1353	**0.0676**	0.6613	0.1208	0.5093	0.0985	0.1208
Mean	0.6873	1.4450	0.98242	0.7420	1.0030	0.6859	0.5622	**0.5275**

Table 1: MSE of the MP, RW, SARIMA, TS, OLS, BM, PLS, and RR methods with different updating months in the testing sample. The minimal values are marked by bold.

While the point forecast results are presented in Table 1, Table 2 shows the optimal λ by minimizing the MSE criterion in the validation set on a dense grid $\{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 3, 5, 7, 10, 15, 50, 10^2, 10^3, 10^4, 10^5, 10^6\}$.

λ	Update month									
	Mar-Dec	Apr-Dec	May-Dec	Jun-Dec	Jul-Dec	Aug-Dec	Sep-Dec	Oct-Dec	Nov-Dec	Dec
PLS	0.1	1	10	50	10	3	50	100	100	1
RR	3	3	7	7	7	3	7	7	3	10^{-6}

Table 2: Optimal tuning parameters used in the PLS and RR methods.

In order for the SVD algorithm used in our updating methods to work, we must have at least two new observations. Thus, it is not possible to update forecasts for Feb-Dec.

6 Discussions and Conclusions

Our approach to forecasting El Niño indices treats the historical data as a high-dimensional vector time series. It has been shown that PCA effectively reduces dimensionality and minimizes the L_2 approximation error. Since principal component scores are uncorrelated, we used an exponential smoothing method to forecast scores, from which one or multi-steps-ahead forecasts are obtained.

We proposed four dynamic updating methods and showed their point forecast accuracy improvement. Although the penalty term of the PLS method is taken to be $q = 2$ in this paper, it is also of common interest to examine other penalty functions, such as $q = 1$. When $q = 1$, the PLS method corresponds to lasso, which is not only a shrinkage method but also a variable selection method. However, the difficulty is that there is no closed form expression for the regression coefficients.

References

Aneiros-Pérez, G. & Vieu, P. (2008), 'Nonparametric time series prediction: A semi-functional partial linear modeling', *Journal of Multivariate Analysis* **99**(5), 834–857.

Bosq, D. (2000), *Linear processes in function spaces: theory and applications*, Springer, Berlin.

Box, G. E. P., Jenkins, G. M. & Reinsel, G. C. (2008), *Time series analysis: forecasting and control*, 4 edn, Wiley, Hoboken, New Jersey.

Dioses, T., & Dávalos, R. & Zuzunaga, J. (2002), 'El Niño 1982-1983 and 1997-1998: Effects on Peruvian Jack Mackerel and Peruvian Chub Mackerel', *Investigaciones marinas* **30**, 185–187.

Frank, I. E. & Friedman, J. H. (1993), 'A statistical view of some chemometrics regression tools', *Technometrics* **35**(2), 109–135.

Hyndman, R. J. & Booth, H. (2008), 'Stochastic population forecasts using functional data models for mortality, fertility and migration', *International Journal of Forecasting* **24**(3), 323–342.

Hyndman, R. J. & Khandakar, Y. (2008), 'Automatic time series forecasting: the forecast package for R', *Journal of Statistical Software* **27**(3).

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

Hyndman, R. J. & Shang, H. L. (2008), Bagplots, boxplots, and outlier detection for functional data, *in* S. Dabo-Niang & F. Ferraty, eds, '*Functional and Operatorial Statistics*', Springer, Heidelberg, pp. 201–207.

Hyndman, R. J. & Ullah, M. S. (2007), 'Robust forecasting of mortality and fertility rates: A functional data approach', *Computational Statistics & Data Analysis* **51**(10), 4942–4956.