Gabriella Piscopo (Italy)

Italian deposits time series forecasting via functional data analysis

Abstract

In this paper we develop a Functional Data Model for forecasting Italian Deposits Time Series. Bank deposits play an important role in ensuring the banks borrowing capacity and for this reason its correct modeling and forecasting represent an interesting task for policy makers. As it is well known, deposit series are affected by seasonality. In the Central Banks and other research institutions the standard procedure applied to this kind of monetary time series is to operate a preliminary seasonal adjustment in order to filter out typical calendar effect and within-year fluctuations. We assume a different starting point in modeling and forecasting seasonal time series, taking into account how the seasonality evolves across the years and trying to incorporate this feature in the model via functional data analysis. We utilize the Phase Plane Plot in order to show the evolution of the seasonality of the Italian Deposits from 1998 to 2008, working on a monthly time series and producing different plots for each year. We fit the data on the historical values using principal component techniques and construct forecast intervals projecting the model components with ARIMA process. The empirical results are presented using a range of graphical analysis.

Introduction

Monetary aggregates represent important information variables for monetary policy-making of Central Banks. Many monetary time series are affected by several seasonal movements that occur during particular periods of the year, like public holidays, tax payment days, school year beginning and so on. Consequently, many statistics are subjected to seasonal adjustment in order to filter out usual seasonal fluctuations and typical calendar effects. In this paper we analyze the fluctuations of bank deposits, which play an important role in ensuring the ongoing banks borrowing capacity. Deposit demand can be influenced by macroeconomic factors, like business cycles and interest differentials with other countries, and by microeconomic factors, like the perceived riskiness of individual banks, liquidity buffers, interest margins and consumption cycles (see Finger and Hesse, 2009). Both aspects have influence on the long run of the time series as well as on the seasonal fluctuations. Despite the diffuse procedure of seasonal adjustments (see Cividini, 1989; Dossé, 1996; Central Bank, 2000; Silvestrini, 2009), we propose a different starting point in the modeling of deposit series, considering seasonality and its evolutions across the years. In this regard, we believe it is interesting to study how the seasonal cycle evolves, in particular in presence of exceptional events, like financial crisis. To pursue this aim, we suggest to apply functional data analysis to deposit time series. As far as it is within our knowledge, this application of functional data procedure has not been previously proposed. The statistical foundations of this work derive essentially from dynamic time series models, nonparametric smoothing techniques and principal component methodology. We do not use any observable variables which could have effects on the deposits, because we do not want to investigate the influence of macroeconomic or microeconomic factors on the deposit demand, but we fit the model on the basis of the deposits historical values, using Singular Value Decomposition. This is the first step of the approach proposed, which principal aim is to forecast deposit series. The statistical tools used are based on a huge literature developed in the last decades and enriched of interesting contributions in the last years. A classical approach to the time series modeling and forecasting consists in using the well known ARIMA process (see Box & Jenkins, 1976) and SARIMA specification for seasonal time series. The problem of time series with high variation and outlier data has been handled with the suggestion of opportune use of smoothing techniques. Simple exponential smoothing has been developed in the 1950s; Holt's linear method (see Holt, 1957) is an extension of simple exponential forecasting that allows a locally linear trend to be extrapolated. For seasonal data the Holts-Winters method has been introduced by Holt (1957); more recent contributions for time dependent data have been proposed in the 1990s (see Härdle W., Vieu P., 1992; Hart J.D., 1996). From the 1990s, the new paradigm of functional data analysis (see Ramsey and Silverman, 1997) has been used for non-parametric modeling and forecasting in a different subject areas, including the contribution in macroeconomic fields. Recently, Aneiros-Pérez and Vieu (2008) have presented an application of functional data analysis to forecast seasonal time series.

The paper is organized as follows: in section 1, we describe the model; section 2 shows the empirical investigation on the Italian deposit time series; section 3 is dedicated to the fitting and forecasting of deposits series via functional data analysis; concluding remarks are offered in the final section.

1. Functional model for seasonal time series

Let $\{Y_\omega, \omega \in [0, \infty)\}$ be a seasonal univariate time series which has been observed at N equispaced point in the time. Given the observed series $Y_1, Y_2, \ldots, Y_N$ the aim is to predict a future value Y_{N+l}, for some integer $l > 1$. The classical method to forecast is to consider an autoregressive process with a parametric approach. The problem can be also treated via non-parametric approach using functional data techniques (see Ramsey and Silverman, 1997). Aneiros-Pérez & Vieu (2008) assume that N can be written as $N = Tp$, where T is the number of trajectories considered and p is the dimensionality of each trajectory, so that the univariate series can be decomposed into T trajectories composed by p observations.

Let $y_t(x)$ be the variable observed in the point x of the trajectory t. We have discrete observations of the time series but reasonably a continuous function describes the evolution of the rate in the continuous time. This function can be approximate by an underlying smooth function $f_t(x)$ that we are observing with errors. Thus, we deal with the functional time series

$$\{x, y_t(x)\}, t = 1, \ldots, T, x = 1, \ldots, p, \text{ where}$$

$$y_t(x) = f_t(x) + \sigma_t(x)\varepsilon_{t,x}, \tag{1}$$

with $\varepsilon_{t,x}$ an iid standard normal random variable and $\sigma_t(x)$ allows for the amount of noise to vary with x. The dataset is smoothed for each t; using a non-parametric smoothing we estimate for each t the functions $f_t(x)$ from $\{x, y_t(x)\}$ for $i = 1, \ldots, p$. We can calculate the derivatives of the curves in order to investigate the velocity and the acceleration of the movements of the variable we are interested in.

The second step of our analysis consists in fitting the data via a basis function expansion using principal component techniques:

$$y_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k} \varphi_k(x) + e_t(x), \tag{2}$$

where $\mu(x)$ is a measure of location of $y_t(x)$, $\{\varphi_k(x)\}$ is a set of orthonormal basis functions and $e_t(x) \sim N(0, \text{var}(x))$. The error term $e_t(x)$, given by the difference between the observed data and the fitted curves from the model, is the modeling error. This basis set provides fit to the estimated curves and gives coefficients that are uncorrelated, simplifying the forecast process. In order to forecast $y_t(x)$, univariate time series models are fitted to each coefficient $\{\beta_{t,k}\}$, $k = 1, \ldots, K$ via singular value decomposition. Using the fitted series $\{\hat{\beta}_{t,k}\}$ the coefficients $\{\beta_{t,k}\}$ $k = 1, \ldots, K$ are forecasted for $t = n+1, \ldots, n+h$ using ARIMA models, structural models (Harvey, 1989) or exponential models (Hyndman et al., 2002). Finally, the previously obtained forecasted coefficients are implemented to get the $f_t(x)$ as in formula (2) and $y_t(x)$ is projected from (1). The estimated variances of error terms in (2) and (1) are used to compute prediction intervals for the forecast.
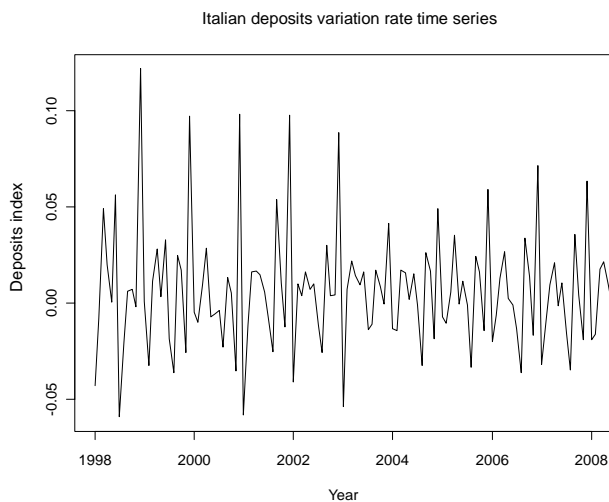
2. Analysis of Italian deposits time series

In the following application we consider the Italian deposits monthly time series from 1998 to 2008; the data can be downloaded from the website of the Bank of Italy (www.bancaditalia.it) where they are collected in table TSC20200. The series is plotted in Figure 1.
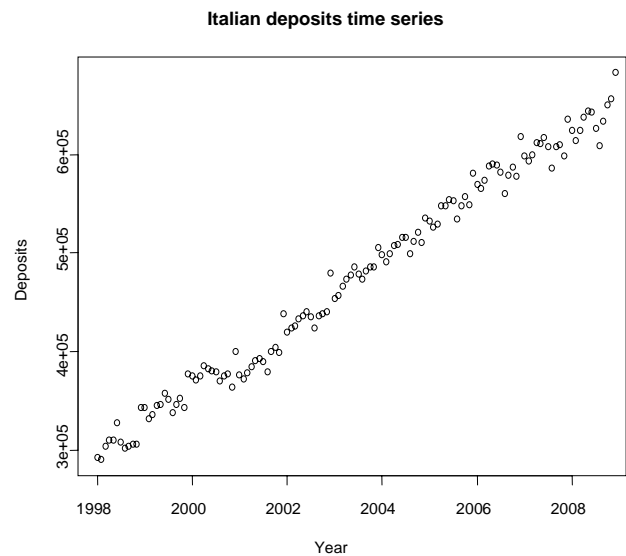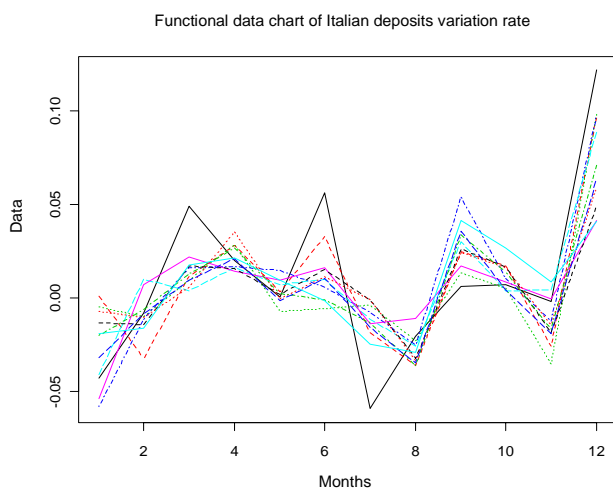

Fig. 1. The Italian deposits monthly time series

The data show seasonality. In Figure 2 the variation rate of the deposits is plotted: it is a univariate series with N=12*11=132 observations. Following Aneiros-Pérez & Vieu (2008), we decompose the series into 11 trajectories with 12 monthly observations. Let $y_t(x)$ be the deposits observed in

the month x of the trajectory/year t; Figure 3 shows the decomposition of the series plotted in Figure 2 into 11 trajectories.

Italian deposits variation rate time series

Fig. 2. The variation rate of the Italian deposits

Functional data chart of Italian deposits variation rate

Fig. 3. The decomposition of variation rate of Italian deposits monthly series

According to the paradigm of Functional Data Analysis, the economic force generating changes in the deposits can be represented by a curve with a certain number of derivatives. We construct this curve $h(t)$ with a B-spline smoothing (see Ramsey and Silverman, 2002) of order 8, considering a number of knots equal to the number of the observations. We want to study how the within-year oscillations change across the years and for this purpose we analyze the first and the second derivatives of the curve. The smoothing method used is designed to give a good impression of the velocity and acceleration of the variable; a data driven technique guides the choice of the model, in order to capture important features in the data and not underestimate peak values or overestimate low values. A useful graphical tool is the so called Phase Plane Plot, i.e. a

plot of the acceleration against the velocity. The graph shows energy which guides the process oscillating between two states: potential and kinetic. Kinetic energy is maximized when the second derivative is zero; potential energy is maximized when velocity is zero. After a period of intense accumulation of deposits or a period of crisis we may see that both potential and kinetic energy are low and the phase-plane curve is close to zero. When we analyze a Phase Plane Plot, we have to look for the cycles, the size of radius of each cycle, the location of the center of the cycles and the changes in the shapes of the cycles across the years. In particular, the larger the radius size is, the more energy is transferred in the process. In the following figures, we show for each year the Phase Plane Plot; visible differences between the graphs give evidence of the evolution of the seasonality. If we look at Phase Plan Plot of the 1998, we note two cycles, the first one is from January to March, and the second one is from March to June and then a period of relative stability. At the beginning of the first cycle the velocity is positive and the acceleration is very high: the deposits increase at an exponential rate. However, the acceleration decreases in the mid of January, at certain point it becomes negative and after a further decrement the process enters the state of maximal potential energy. In this case, the energy developed is negative, the first derivate becomes negative and the deposits decrease. At the end of January the process is in the state of kinetic energy: the deposits continue to fall. Between February and March potential energy is developed and the deposits return to grow. Between March and June the second cycle takes place, with less intensity; as expected, deposits show a negative velocity in April, when the Easter holidays and tax payments occur, and in June, during the summer holiday. The cycle is followed by a period of relative stability. We can also find this period during 1999 and 2000. From 2001, the pattern of the process is modified and the number of cycles rises. In 2003, the velocity is positive until June and the curve lies on the right part of the plot; between June and August and between October and November the velocity appears negative and the deposits fall: these periods coincide with the summer holidays and the school year beginning, when the withdrawals increase. A similar feature is tracked in 2004, 2005, 2006 and 2007: as expected, the deposits fall between June and July and between October and November. Finally, in September 2008 when financial crisis exploded instability enhanced.
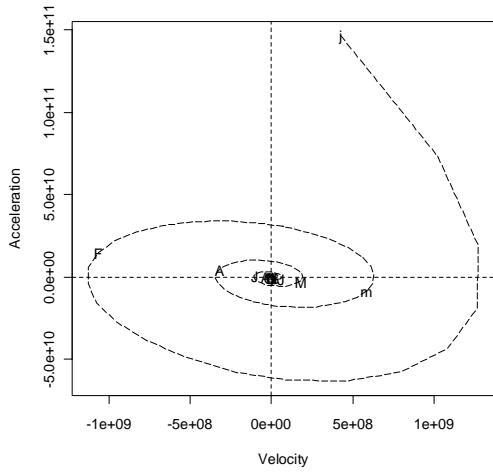
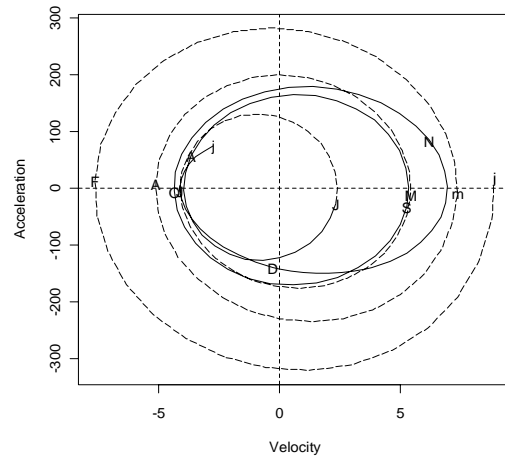**Fig. 4. The phase plane plot of the Italian deposits for 1998**

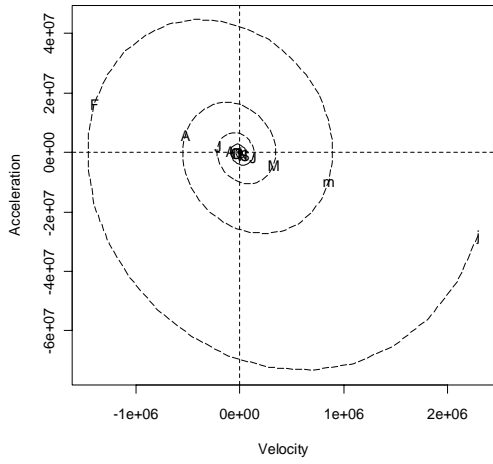**Fig. 5. The phase plane plot of the Italian deposits for 1999**

Fig. 6. The phase plane plot of the Italian deposits for 2000

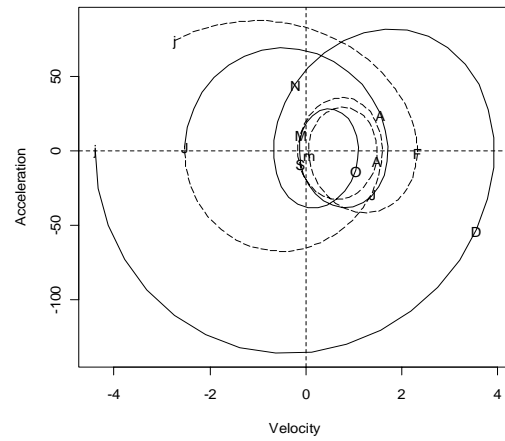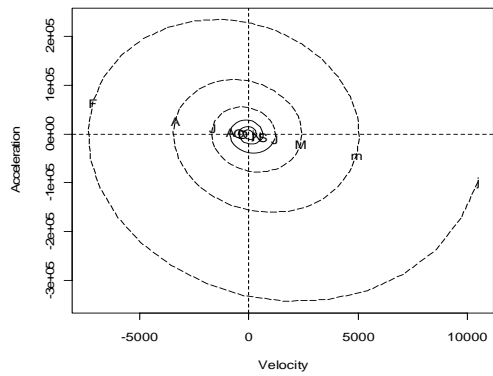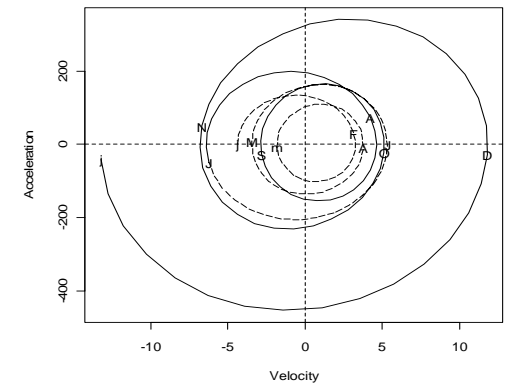**Fig. 7. The phase plane plot of the Italian deposits for 2001**

Fig. 8. The phase plane plot of the Italian deposits for 2002

Fig. 9. The phase plane plot of the Italian deposits for 2003

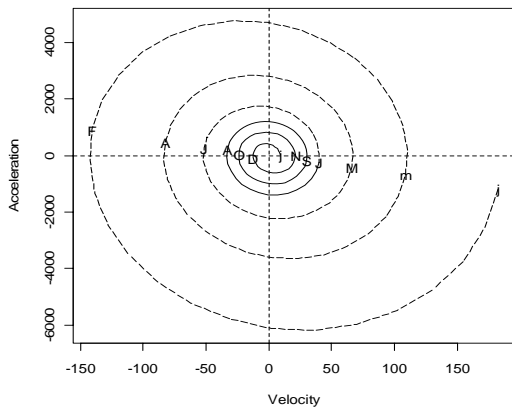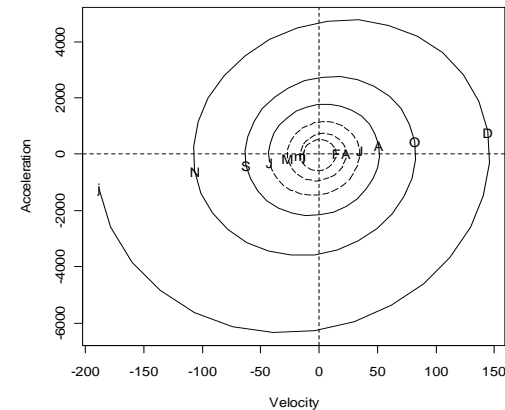**Fig. 10. The phase plane plot of the Italian deposits for 2004**

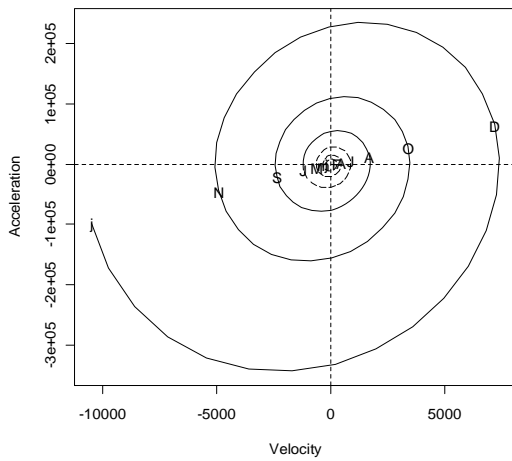**Fig. 11. The phase plane plot of the Italian deposits for 2005**

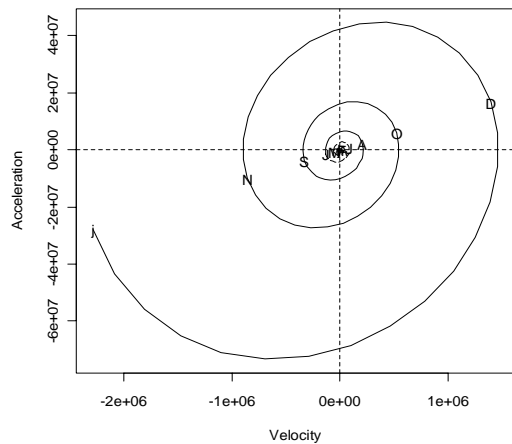**Fig. 12. The phase plane plot of the Italian deposits for 2006**

Fig. 13. The phase plane plot of the Italian deposits for 2007
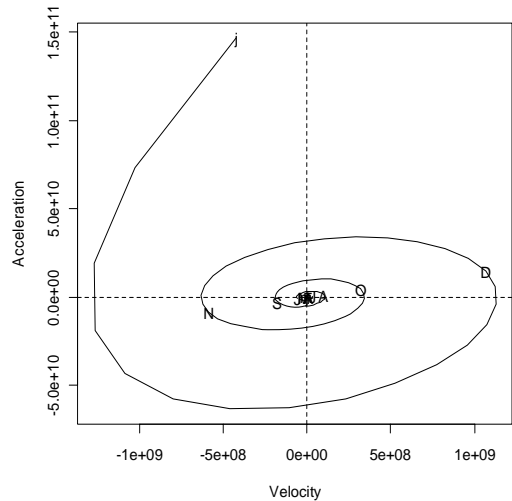

Fig. 14. The phase plane plot of the Italian deposits for 2008

## 3. Functional data fitting and forecast of the deposit series

In the previous section we have shown how the deposits variations change across the years. Now we want to produce forecasts for the next year. The problem can be treated via non-parametric methods using functional data techniques. Let $y_t(x)$ be the deposits observed in the month x of the

trajectory/year t. The observed time series $\{Y_1, Y_2, \ldots, Y_{132}\}$ can be divided into 11 paths of length 12 in the following sets: $y_t = \{Y_\omega, \omega \in (p(t-1), pt]\}, t = 1, \ldots, 11$. The aim is to forecast future deposits value $y_{T+h}, h > 0$ from the observed data. To pursue this scope, we apply a non-parametric method: via principal component analysis we decompose the (12×11) matrix $(Y = [y_1, \ldots, y_{11}])$ into a number of principal components and associated coefficients using the Singular Value Decomposition:

$$Y = \hat{\mu} + \hat{\varphi}_K' \hat{\beta}_1 + \ldots + \hat{\varphi}_K' \hat{\beta}_K + \hat{\varepsilon} ,$$

where $\hat{\mu} = [\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_{12}]'$ is the mean vector, $\hat{\varphi}_1, \hat{\varphi}_2, \ldots, \hat{\varphi}_K$ are the estimated principal components ($\hat{\varphi}_k = [\hat{\varphi}_{1,k}, \ldots, \hat{\varphi}_{12,k}]$); $\hat{\beta}_1, \ldots, \hat{\beta}_k$ are uncorrelated basis functions ($\hat{\beta}_k = [\hat{\beta}_{1,k}, \ldots, \hat{\beta}_{11,k}]'$); $\hat{\varepsilon}$ is the zero mean (12×11) residual matrix. Figure 15 shows fitted basis functions and coefficients for an expansion of order 8. We highlight that increasing the order of the model the basis functions estimated in the previous model do not change and only other functions are considered; this is due to the methodology of Singular Value Decomposition. The expansion order is chosen according to a data driven technique; we have increased the number of basis functions until residuals from the fitting have shown an irregular pattern. In the first line of Figure 15, the first graph represents $\hat{\mu}$ across the months and the other graphs show the fitted basis functions $\hat{\beta}_k, k = 1, \ldots 8$; in the second line the graph represents the fitted coefficients $\hat{\varphi}_k, k = 1, \ldots, 8$. As it is clear from Figure 15, the basis functions model different movements in the deposits across the months. For this reason, the model is able to explain the within-year variations. The mean term is a month-specific component independent of the years, the coefficients are year-varying parameters. In particular, the first coefficient represents the trend in the deposit series and the first basis function explains how deposits change in each month when the general level in the deposits moves. The second basis function mainly models the first months of the year, while the third one models the last months. The other functions are more complex and capture the differences between all months.
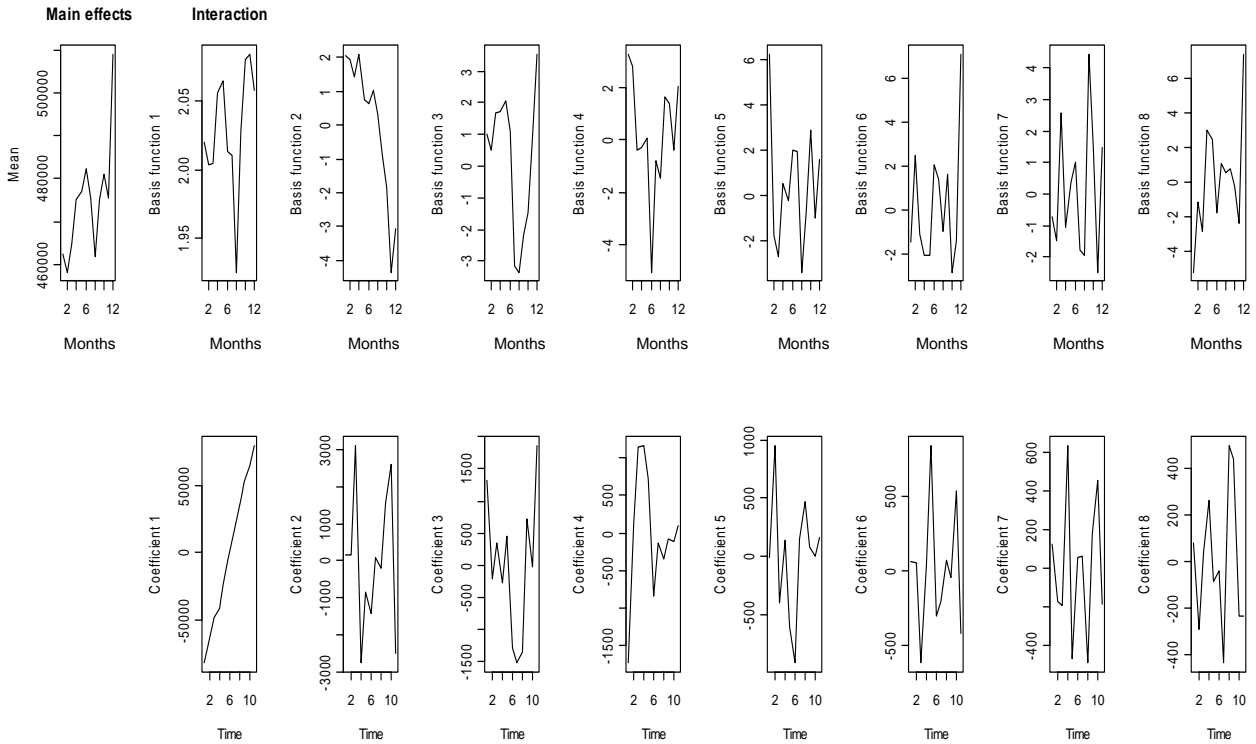
Fig. 15. The fitted coefficients and basis functions

In the analysis conducted we have assumed a different point of view with respect to traditional seasonal adjustment methods for time series. One of these is the widely used decomposition of time series by loess (cf. Cleveland et al., 1990). In the following, we present the classical decomposition in order to compare the results obtained with respect to functional analysis and reflect them upon the two different approaches. Figure 16 shows the decomposition of Italian deposits time series into trend and seasonal effect.
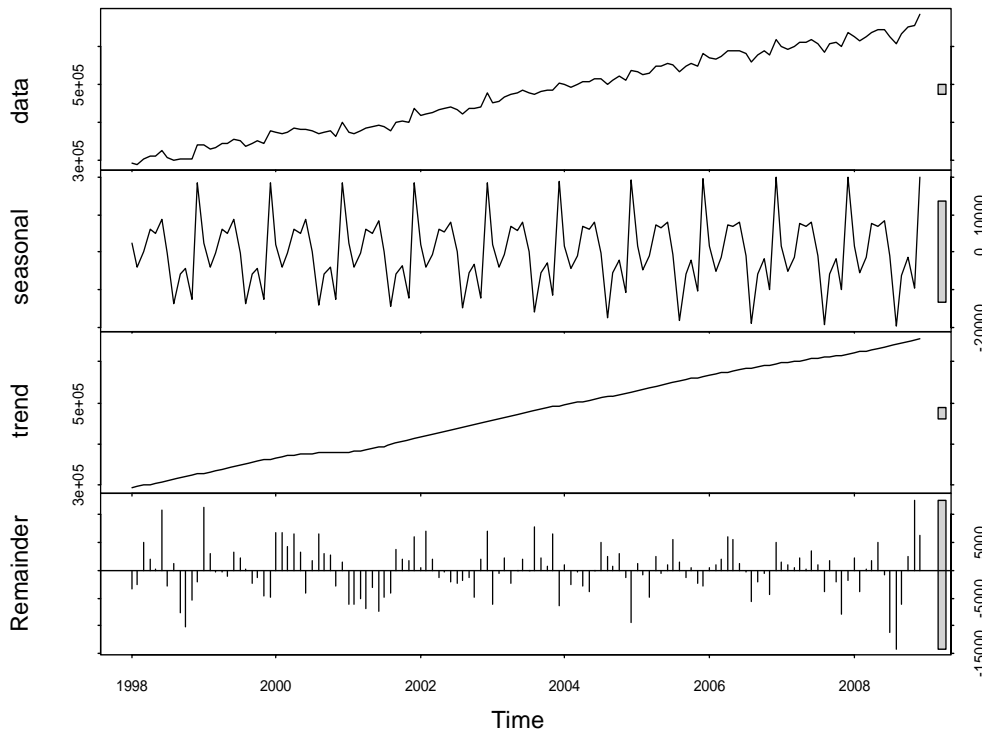
Fig. 16. The decomposition of Italian deposits time series

Using this approach movements in seasonality are not pointed out, unlike what happens applying functional data analysis. Consequently, in the former case the model neglects some important data features and this impacts residuals. Figures 17 and 18 show residuals from the models fitted with classical decomposition and functional technique: in the latter graph, the dispersion around zero is smaller.
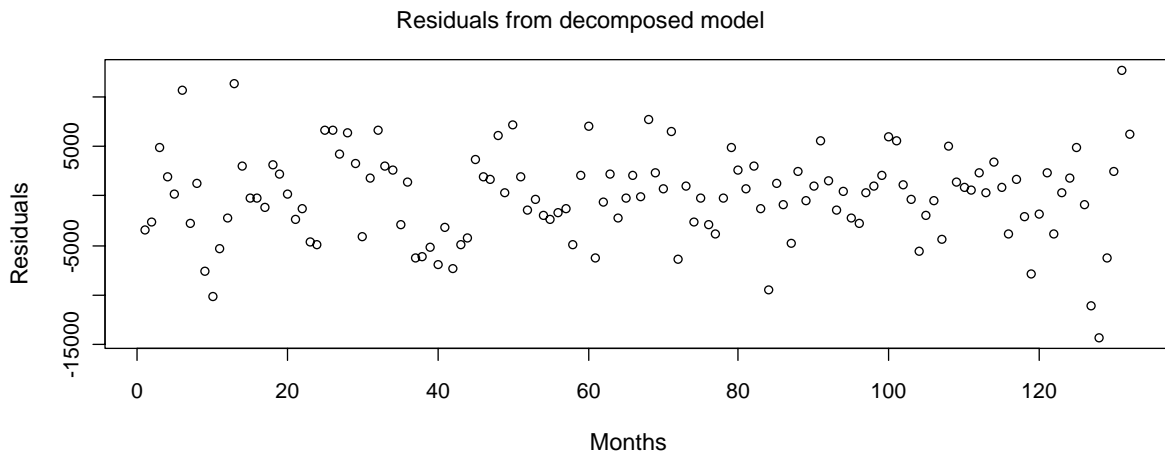
17

Residuals from decomposed model

Fig. 17. Residuals of the Italian deposits fitted series through decomposition by loess

Residuals from functional model
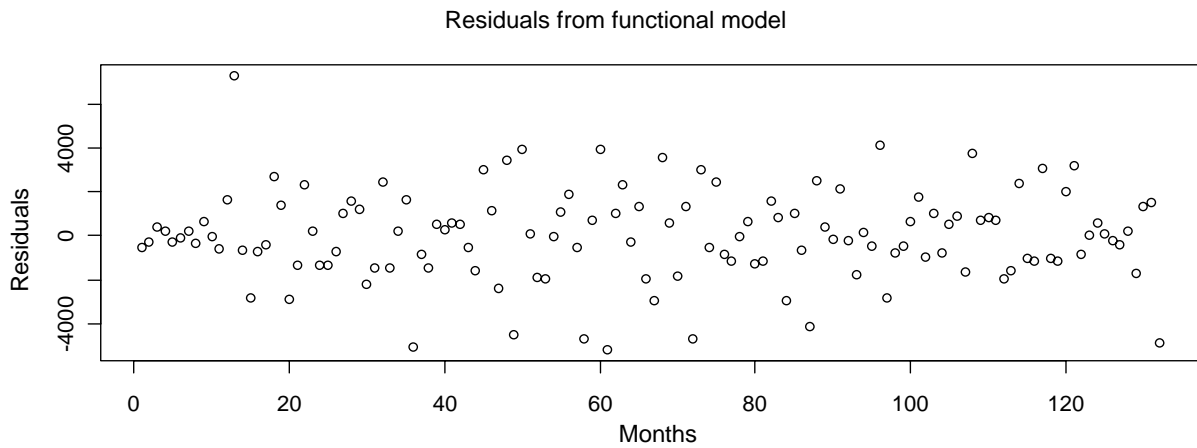
Fig. 18. Residuals of the Italian deposits fitted series through functional data analysis
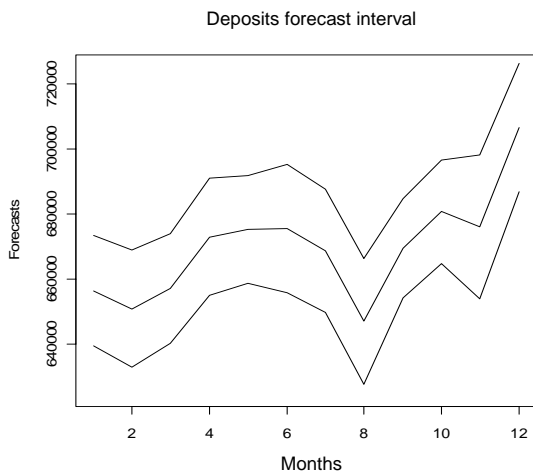
Deposits forecast interval

Fig. 19. The forecast interval for the deposit time series

The last step of our work consists in producing forecast intervals for the deposits during the next year starting from the fitted coeffients and basis function of the functional model. Then, conditioning on the observed data $\Psi = \{y_t(x_i); t = 1, \cdots, T; i = 1, \ldots, p\}$ and on the set of basis functions Φ, we obtain the h-step forecasts

$$\widetilde{y}_{T,h}(x_i) = E[y_{T+h}(x_i)|\Psi, \Phi] = \hat{\mu}(x_i) + \sum_{k=1}^{K} \widetilde{\beta}_{T,k,h} \hat{\varphi}_k(x_i),$$

where $\widetilde{\beta}_{T,k,h}$ denotes the h-step ahead forecast of

$\beta_{T+h,k}$ obtained using the estimated time series $\{\hat{\beta}_{t,k}, t = 1, \ldots, T\}$ and projecting them with opportune ARIMA process. Figure 19 shows the results with a level of confidence equal to 95%.

Concluding remarks

In this paper we have developed a Functional Data Model to analyze and forecast Italian Deposits Time Series. As Ramsay and Silverman (2002) write, *"the aim of the analysis of functional data are...to formulate the problem at hand in a way amenable to statistical thinking and analysis, to develop ways of presenting the data that highlight interesting and important features; to investigate variability as well as mean characteristics..."*. In this regard, we have assumed a different point of view with respect to standard procedures implemented by Central Banks and other research institutions, which apply preliminary seasonal adjustment in order to filter out typical calendar effect and within-year fluctuations of deposit series. Our proposal does not set against the current modus operandi, but can be used as a complementary tool, useful for highlighting important features of data, like seasonality and its evolution across the years. The results shown by the

ready-understanding Phase Plane Plots can be interpreted in the light of events which have impact on the economy, like political occurrences or the recent financial crisis. The variation of the seasonality is captured in fitting and forecasting dealing with basis expansions of time series. Further works can investigate the forecast accuracy of the model. Wide space for the future research is offered by the implementation of functional data techniques able to produce estimations robust to the outliers data.

References

1. Aneiros-Pérez G., Vieu P. (2008). Nonparametric time series prediction: a semi-functional partial linear modeling. *Journal of Multivariate Analysis*, 99(5), pp. 834-857.
2. Box G.E.P., Jenkins G.M. (1976). Time Series Analysis: Forecasting and Control. Holden-Day.
3. Cleveland R.B., Cleveland W.S., McRae J.E., Terpenning I. (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics*, 6, pp. 3-73.
4. Cividini A., Cottarelli C. (1989). Procedure di Destagionalizzazione dei Depositi Bancari Mensili in Italia. Temi di discussione n. 112, Banca d'Italia.
5. Dossè J.E., Planas C. (1996). *Revisions* in Seasonal Adjustement Methods: An Empirical Comparison. Working Group Document D3/SA/07, EUROSTAT.
6. Finger H., Hesse H. (2009). Lebanon – Determinants of Commercial Bank Deposits in a Regional Financial Center. International Monetary Fund, WP/09/195.
7. Härdle W., Vieu P., (1992). Kernel regression smoothing of time series. *Journal of the Time Series Analisys*, 13, pp. 209-232.
8. Hart J.D. (1996). Some automated methods of smoothing time-dependent data. *Journal of Nonparametric Statistics*, 6, pp. 115-142.
9. Holt C.C. (1957). Forecasting Trends and seasonal by exponentially weighted averages, O.N.R., Memorandum 52/1957, Carnegie Institute of Technology. *Reprinted in International Journal of Forecasting* (2004), 20, pp. 5-10.
10. Hyndman R.J. (2009). Nonparametric time series forecasting with dynamic updating. 18th World IMAS/MODSIM Congress, Cairns, Australia.
11. Ramsey J.O., Silverman B.W. (1997). Functional Data Analysis. New York: Springer-Verlag
12. Ramsey J.O., Silverman B.W. (2002). Applied Functional Data Analysis. New York: Springer-Verlag.
13. Silvestrini A. (2009). *La* destagionalizzazione dei depositi e dei prestiti bancari. *Questioni di Economia e Finanza*, n. 42, Banca d'Italia.