A wavelet threshold denoising procedure for multimodel predictions: An application to economic time series: FENGA

Article *in* Statistical Analysis and Data Mining · July 2017

DOI: 10.1002/sam.11351

CITATIONS
2

READS
113

1 author:

Livio Fenga
University of California, San Diego
14 PUBLICATIONS **73** CITATIONS

SEE PROFILE

A Wavelet Threshold Denoising Procedure for Multimodel Predictions: an Application to Economic Time Series

Livio Fenga

ISTAT, National Institute of Statistics
fenga@istat.it

Abstract. Noise–affected economic time series, realizations of stochastic processes exhibiting complex and possibly non-linear dynamics, are dealt with. This is often the case of time series found in economics, which notoriously suffer from problems such as low signal-to-noise ratios, asymmetric cycles and multi–regimes patterns. In such a framework, even sophisticated statistical models might generate suboptimal predictions, whose quality can further deteriorate unless time consuming updating or deeper model revision procedures are carried out on a regular basis. However, when the models' outcomes are expected to be disseminated in timeliness manner (as in the case of Central Banks or national statistical offices), their modification might not be a viable solution, due to time constraints. On the other hand, if the application of simpler linear models usually entails relatively easier tuning–up procedures, this would come at the expenses of the quality of the predictions yielded. A mixed, self-tuning forecasting method is therefore proposed. This is an automatic, two–stage procedure, able to generate predictions by exploiting the denoising capabilities provided by the wavelet theory in conjunction with a compounded forecasting generator. Its out–of–sample performances are evaluated through an empirical study carried out on macroeconomic time series.

Keywords: SARIMA models; SETAR models; Time series forecast; Wavelet

1 Introduction

In many fields of theoretical and applied research, mathematical models aimed at predicting the future values taken by a random variable are routinely employed. Forecasting is a widespread and vital activity in many fields of research – such as engineering, medicine, climatology and economics – which requires theoretical and computational efforts in an amount usually proportional to the degree of uncertainty of the phenomenon under investigation. Economics is certainly a major field of application of prediction methods. Here, strong uncertainty is often paired with the need of accurate predictions, generally released in timeliness fashion and yielded by ad hoc models, i.e. carefully designed according to

the purposes envisioned, such as of monitoring key indicators, implementing or refining decision processes, gaining precious insights about future or on-going strategies, e.g. in terms of inflation rate, industrial production index, GNP, unemployment rate, welfare and so forth. That economic prediction is an activity in a constant rise and good methods are always in demand, is explainable, at least in part, by two main phenomena: World increasing level of complexity and data proliferation. Regarding the first aspect, as the World increases its level of complexity, virtually all the processes related to its manifestations embody, to a different extent, such a feature, which might appear in several forms, such as noise, irregular or even erratic behaviors and many other types of potentially disturbing factors. This is particularly true having in mind events – often catastrophic – that have characterized the first years of the 21^{st} century. Economic crises (speculative bubble), wars (Iraq, Afghanistan, Ukraine, Isis), terrorist attacks (Twin Towers), health emergencies (the spread of Ebola virus) and several others, have introduced, with different intensities, strong elements of uncertainty – such as spillovers and volatility – in the economic system. As a result of that, models successfully applied previously, need, more and more often, to be revised, redesigned or even replaced with more appropriate alternatives, i.e. able to account for such interventions in order to yield more reliable forecasts. On the other hand, the second aspect reflects today's availability of bigger than ever quality information–sets, related to a growing number of economic variables. In this regard, it is also remarkable the extent their sampling frequencies have been increased to, mainly thanks to the progress achieved in computer sciences and engineering (e.g. web–based data capture techniques, satellite imaging methodologies, and so on). In fact, nowadays many time series are recorded at frequency rates unattainable even in the recent past, so that sampling frequencies of few minutes, as in the case of the electric power demand or expressed in a tick-by-tick fashion, as in the area of financial markets, are not uncommon. The latter is actually the case where the time interval between human (or machine-driven) action into the real world and the generation of the related measures – often in the form of ready-to-use statistical indicators – virtually collapses. Because of this scenario – characterized by a spectacularly huge amount of data, often available at virtually no cost for the end user – building, tuning up and maintain a complex, non-linear prediction model is surely a challenging task, which can become quickly cumbersome when time constraints are in place. On the other hand, simpler linear models are in general easier to manage but likely unable to fully account for the complicated patterns more and more often exhibited by macroeconomic time series. Despite that, much statistical analysis of time series, is still done under the linear paradigm where the use of denoising algorithms appears to be quite limited. Such a framework – outside trivial cases and ad hoc lab–controlled experiments – hardly ever do possess features compatible with the dynamics encountered in real-world data analysis setups. Capturing such features is the aim of the proposed forecast procedure, designed to minimize human interaction, which combines the well established wavelet threshold denoising approach, with a mixing–forecasting procedure, based on the combina-

tion of the predictions delivered by a linear model of the class SARIMA (Seasonal Autoregressive Moving Average) and SETAR (Self Exciting Threshold Autoregressive). The framework presented here is quite flexible, as it is able to capture multi-regimes type behaviors, symmetric and asymmetric cyclical components as well as linear features of the series at hand. It is also able to account for and balance the relative importance of its linear and non-linear features, through an optimal weighted combination of the predicted values yielded by the two models considered. Such predictions are extracted from a filtered version of the original time series and combined according to the method proposed by [1].

2 Signal denoising and prediction procedures

In their raw format, most of the real-world signals are available in time–domain and, as such, have a related physical representation of the type time–amplitude. In this regard, throughout the present paper, the time series of interest is intended to be a real–valued, uniformly sampled, sequence of data points of length T, denoted as:

$$x_t := \left\{ \left(x_t \right)_{t \in \mathbb{Z}^+}^{T} \right\}, \tag{1}$$

whereas its predicted values at horizon h are formalized as follows:

$$x_t(h) = \left\{ \left(x_t \right)_{t \in \mathbb{Z}^+}^{T+h}; \qquad h = 1, 2, ..., H \right\}. \tag{2}$$

This is perhaps the most intuitive representation, given the analogical basis it is grounded upon, but not necessarily the optimal one for any sort of applications or analysis. In the present paper, both the domains are employed in a multistage and iterative fashion. In more details, the frequency information conveyed by the data are used to denoise the time series – through a wavelet thresholding-type procedure – whereas the time domain part is in charge of generating optimally weighted predictions – minimizing a suitable forecast quadratic loss function.

In more details, the presented procedure – called FOR-WARD (short for FORecast via Wavelet Automatic Real-valued time series Denoiser) to emphasize its self-adjusting forecasting and denoising capabilities, is based on an iterative procedure aimed at finding the optimal wavelet filter-denoising thresholds, conditioned to the set of predictions yielded by the two models considered (which are formalized in Eqn. 7 and 10 *and* the optimal weighting parameters used to combine these predictions. In other words, the original time series is iteratively filtered according to a set of thresholds and processed independently by optimal (in a sense explained below) SARIMA and SETAR models until a suitable quadratic loss function is minimized.

2.1 Wavelet–based denoiser

As suggested by their name, wavelets – formalized in Eqn. 3 and 4 – are well localized functions whose amplitude grows from zero, reaches a maximum, and then decreases back to zero again. They play a key role in FOR–WARD procedure, as it uses wavelet expansions to generate smoother versions of the original signal. In the sequel, only the function Hilbert space $L^2(\mathbb{R})$ of all the squared integrable functions is considered, i.e. $f \in L^2(\mathbb{R})$ iff: $\int_{\infty}^{+\infty} |f(x)|^2 dx < \infty$, satisfying a function, say $\psi(x)$, verifying

$$f(x) = \sum_{k\in\mathbb{Z}} c_k \phi_{0,k}(x) + \sum_{j<J,k\in\mathbb{Z}} d_{j,k}\psi_{j,k}(x),$$

where

$$c_k = \int_{\mathbb{R}} f(x)\phi_{0,k}(x)dx \qquad (3) \qquad d_{j,k} = \int_{\mathbb{R}} f(x)\psi_{j,k}(x)dx, \qquad (4)$$

being J the maximum number of allowable resolutions, given the available sample size. Here, each set of coefficients, usually referred to as crystal, is linked to a spacial scale j, whereas every single coefficient k, called atom, accounts for a particular location. In practice, the wavelet coefficients $\{d_{jk};\quad j = 1, 2, ..., J\}$ (Eqn. 4) account for and represent progressively finer and finer details whereas smooth dynamics at the coarsest scale are captured by the crystal $\{c_k\}$ (Eqn.3). The function $\psi(x)$, called mother wavelet, generate all the wavelet functions $\psi_{j,k}(x)$ through the equation $\psi_{j,k}(x) = 2^{\frac{1}{2}}\psi(2^j x - k)$, so that the set $\{\psi_{j,k}\}$ defines an orthonormal basis for $L^2(\mathbb{R})$. By dilations and translations of $\psi(x)$, the space of functions $\Psi(x)$ is generated, i.e.

$$\Psi = \left\{ \psi_{c,b}(x) = |c|^{-1/2}\psi\frac{x-b}{c},\ \ c, b \in \mathbb{R}, \quad c \neq 0 \right\},$$

with the function ψ satisfying the following conditions: *a)* $\int_{-\infty}^{\infty} \psi(t)dt = 0$; *b)* $\int_{-\infty}^{\infty} |\psi(t)|dt < \infty$; *c)* $\int_{-\infty}^{\infty} \frac{|\hat{\psi}(\xi)|}{|\xi|}dt\xi < \infty$, $\hat{\psi}(\xi)$ being the Fourier transform of $\psi(t)$; *d)* $\int_{-\infty}^{\infty} t^j\psi(t)dt = 0$, $j = 0, 1, ..., r-1$, under the conditions that it must exist at least a $r \geq 1$ and $\int_{-\infty}^{\infty} t^r\psi(t)dt < \infty$. Even though different approaches are available, the proposed method has been implemented with a signal–coefficient transformation procedure of the type Discrete Wavelet Transform (DWT), associated with boundary condition of the type periodic. Basically, the observed time series x_t is modified by artificially introducing an extension of it, so that the unobserved samples $\{x\}_{t\in Z^-}$ are assigned the observed values x_{T-1}, x_{T-2},...,x_0. This method considers the series as it were periodic, and is known as using circular boundary conditions. Now, the wavelet and scale coefficients are given by:

$$d_{j,t} = \frac{1}{2^{j/2}} \sum_{l=0}^{L_j-1} h_{j,l}, x_{t-l \mod N}, \qquad S_{J,t} = \frac{1}{2^{j/2}} \sum_{l=0}^{L_j-1} g_{j,l}, x_{t-l \mod N},$$

where $\{h_{j,l}\}$ and $\{g_{j,l}\}$ are the length L, level j, wavelet and scaling (approximate) filters respectively: the former, of the type band–pass, with nominal pass–band $f \in [\frac{1}{4\ \mu_j}, \frac{1}{2}\mu_j]$, and the latter of the type low-pass, with a nominal pass-band $f \in [0, \frac{1}{4}\mu_j]$, with μ_j denoting the scale.

Statement of the Problem Since real–life data are always supposed to be observed with error, no direct access to the theoretical, uncorrupted realizations $y_t \sim \mathfrak{M}(\cdot)$, with $\mathfrak{M}$ the statistical model followed by the DGP, is possible. In fact, one is only able to observe the realization x_t (Eqn. 1) – i.e. $x_t = y_t + \varepsilon_t$ – which is supposed to be measured with additive, independent noise ε_t. In this case, wavelet theory can be fruitfully employed to achieve a better signal–to–noise ratio, for the benefit of virtually all the stages of the analysis. For example, the investigation of the probabilistic structure of the time series at hand via the Empirical Autocorrelation Functions, the inference methods of the type Maximum Likelihood as well as the model order selection and validation procedures can be, to varying degrees, biased due to the presence of noise components. The consequences are in general not negligible and range from instability of the parameter estimates (resulting in significant amount of model uncertainty), to the selection of the wrong model. The adopted theoretical framework is now illustrated: given the observed data $\{x_1, \dots x_i, \dots x_n\}$, assume the model

$$x_i = f(t_i) + \sigma\varepsilon_i, \tag{5}$$

with $t_i = \frac{1}{n}$, to be measured with additive, independent noise, i.e. the sequence $\{\varepsilon\}_1^n$, with unknown variance. The function f has to be estimated from the available stretch of data $\boldsymbol{x_i}$, according to a suitable quadratic loss function $\mathcal{L}$, so that, defining $\hat{f}$ the estimated function, we have: $\hat{f} = \min_{\mathcal{L}} ||\hat{f} - f||_2$. Defining $x_{j,k}$ as the j–level crystal ($j \subset J$) – with k being the index–identifier for the k–atom – the wavelet-based denoiser enter this setup by considering that Eqn. 5 – in the wavelet domain – under wavelet orthogonality and periodic conditions– reads as

$$x_{j,k} = w_{j,k} + \sigma\eta_{j,k},$$

where $w_{j,k}$ are the noise–free wavelet coefficients and η is the N(0,1) *iid* sequence. Taking advantage of the sparse structure of the crystal (as many 0–coefficients are likely to be embedded in the theoretical pure signal), the problem can now be restated as follows: instead of focusing on the unknown function f, we think in terms of those atoms whose magnitude is above (below) a certain threshold in order to retain them as an integral part of the signal (discard them as an unwanted component). Such a discriminating operation is conducted on those crystals accounting for higher frequencies, as they are usually the most affected by noise, unlike the crystals accounting for the lowest frequencies. Thresholding methods usually are level dependent – i.e. $\lambda = \lambda(j)$ or, in more complex cases, based on both level and atoms – i.e. $\lambda = \lambda(j, k)$. Once estimated, the threshold

can be applied according to different rules; the one denominated hard threshold scheme, expressed as follows,

$$\delta(\lambda) = \begin{cases} 0, & \text{if } d_{j,k} \leq \lambda \\ d_{j,k}, & \text{if } d_{j,k} > \lambda, \end{cases} \tag{6}$$

is the one adopted in the present paper.

2.2 The Prediction Models and Selectors employed

In what follows, the stochastic models considered in the present study are briefly summarized. As popular statistical tools, they have been often employed to capture various types of linear – as in the case of the SARIMA models, popularized by [2] – and non linear dynamics – as in the case of models of the type SETAR, introduced by [3] and further developed by [4]. In particular, the former has been successfully employed in many econometric applications concerning the prediction of different variables, such as electric power load – see, for example, [5], [6], [7] – tourism demand [8] or inflation rate [9]. On the other hand, SETAR models are useful in those circumstances when the non–linear features of the time series at hand are not properly accounted for by pure linear models. This is, for instance, the case when a significant part of the total information is stored in the non linear components of the time series under investigation, such as asymmetric cycles, multimodal jumps or multiple regime dependency. Empirical exploration of their performances may be found in many sources (e.g. [10] and [11]), whereas specific econometric applications can be found in different contexts, such as interest rates ([12], [13]), output growth [14] and stock market ([15], [16]).

With regard to the selection of the "best" model order, it should be emphasized that many identification strategies and selectors are nowadays available, also in the form of ready-to-use fast routines, implemented in many commercial and free of charge software package. Among them, information criteria have gained widespread acceptance by both theoretical statisticians and practitioners. An excellent summary of this criteria can be found in [17] whereas a through discussion, limited to the ARMA case, is available in [18]. In essence, information criteria provide a metric for balancing between goodness of fit and complexity of the model and select an order that minimizes a quantity expressed by the maximized log likelihood plus a penalty term. The Information Criterion of Akaike (AIC), proposed by [19] – probably one of the most well-known and successfully employed order selector – is the criterion applied in the present paper. Based on an estimate of the expected relative entropy (the Kullback - Leibler divergence) contained in an estimated model – i.e. the degree of divergence from the "true" theoretical model – it has the desirable properties of belonging to the class of Shibata efficient criteria and of being designed for problems of the type ∞–dimensional. Unfortunately, as it has been proven by [20], AIC penalty term does not guarantee consistency of the selector so that, by adding more and

more structure to candidate models, it becomes negatively biased and, a result, overfitting will arise with non zero probability. This Criterion can be defined as follows:

$$AIC = -2 \max \log(L(\hat{\theta}|x)) + 2K,$$

where K is the dimension of the model and $(L(\hat{\theta}|x))$ is the log likelihood function. Finally, for the sake of notation convenience, in the sequel SARIMA and SETAR models will be respectively referred to, either in the text and in the formulae, as model a and b and identified through the indicator variable $I = \{a, b\}$.

SARIMA models A generalization of the $ARIMA$ class proposed by [2], $SARIMA$ models have been introduced to model complex dynamics of the type stochastic seasonal. Let x_t be an observed stretch of data with mean μ satisfying (Eqn. 1), $SARIMA$ model can be expressed as:

$$\phi_p(B)\Phi_P(B^S) \bigtriangledown_S^D \bigtriangledown^d (x_t - \mu) = \theta_q(B)\Theta_Q(B^S)\alpha_t, \tag{7}$$

where – denoting with B the backward shift operator, d and D the non-seasonal and seasonal difference operator respectively, defining $\bigtriangledown^d = 1 - B^d$ and $\bigtriangledown^D = 1 - B^D$ – we have $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - .... - \phi_p B^p$, $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - .... - \theta_q B^p$, $\Phi_P(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2S} - .... - \Phi_P B^{PS}$ and $\Theta_Q(B^S) = 1 - \Theta_1 B^S - \Theta_2 B^{2S} - .... - \Theta_q B^{QPS}$. Here, ϕ, θ, Φ, Θ, respectively denote the non-seasonal autoregressive and moving average parameters and the seasonal autoregressive and moving average parameters. Finally α_t is a 0-mean white noise with finite variance. The SARIMA vector of parameters, in the sequel denoted with the Greek letter ${}^a\boldsymbol{\xi}$, is expressed as

$${}^a\boldsymbol{\xi} \equiv (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \Phi_1, \ldots, \Phi_P, \Theta_1, \ldots, \Theta_Q; \mu, \sigma^2),$$

whereas the SARIMA order space is formalized as

$${}^a\boldsymbol{\Xi} \equiv (p, d, q, P, D, Q). \tag{8}$$

Eqn. 7 can be estimate when the stationary and invertibility conditions are met for both the autoregressive and moving average polynomials respectively, that is when the equations $\phi_P(B)\Phi_p(B)$=0 and $\Theta_Q(B)\theta_q(B)$=0 have their root lying outside the unit circle. Generally, $SARIMA(p, d, q)(P, D, Q)^S$ is how Eqn. 7 is abbreviated. Finally, following [21], the AIC for this class of models can be expressed through Eqn 9, i.e.:

$$AIC(p, q, P, Q) = T \ln \hat{\sigma}_\epsilon^2 + 2 \left\{ \frac{T}{T - (d + D)} \right\} + 2(p+q+P+Q) + 2 + T \log 2\pi + T, \tag{9}$$

whose parameters are as above defined.

SETAR models An observed stretch of data, say $(x_{-U+1},,x_{-1},x_0,x_1,,x_T)$, is a realization of a SETAR–type DGP when can be represented as

$$x_t = \varphi_{j0} + \sum_{k=1}^{p_j} \varphi_{jk} y_{t-k} + \varepsilon_{jt} \qquad \gamma_{j-1} < y_{t-d} \leq \gamma_j, \tag{10}$$

where φ_{jk} $(k = 1, ...p_j)$ are the parameters in state j, $(j = 1, 2, ..., l)$, ε_{jt} is a 0-mean variance σ^2 Gaussian White Noise, $-\infty = \gamma_0 < \gamma_1 ... < \gamma_l = +\infty$ are the threshold values and d is the delay parameter. Finally, with U the number of the initial conditions is denoted, i.e. $U = \max(p_j, d); \quad j = 1, 2, .., l$. The vector of parameters of model 10, in the sequel denoted with the Greek letter ${}^b\boldsymbol{\xi}$, is as follows:

$${}^b\boldsymbol{\xi} \equiv (\phi_{10}, \phi_{11}, ..., \phi_{1p_1}, ..., \phi_{l0}, ..., \phi_{lp_l}; \gamma_1, ..., \gamma_l, d, \sigma_1^2, ..., \sigma_{p_l}^2),$$

whereas the SETAR order space is formalized as

$${}^b\boldsymbol{\Xi} \equiv (j, l, d). \tag{11}$$

Model estimation strategies can be carried out under stationary and ergodic conditions of Eqn. 10, finiteness of its second moments and positiveness (everywhere) of its associated density $(x_1, x_2, \cdots, x_T)'$. One of the most common is simply based on the conditional least squares and the minimization of the predictive sum of squared errors obtained as a result.

On the other hand, the identification of the structures for these models is a much more complicated task. In this regard, FOR–WARD uses an objective function for each of the considered horizons $h \in H$ and the filtered version of the time series, so that a sequential exhaustive search on a pre-specified model order space (11) is performed. The final conditional model is then selected via MAICE (Minimum AIC Expectation) procedure, whose optimality has been discussed by [22] and [23] and which will be briefly summarized in Paragraph 2.3. A suitable version of the AIC, called *pooled-AIC* – employed in different applications, e.g. in [24] and [25] – will be used in the empirical study. Its estimation is conditional upon both a specified threshold delay and a fixed threshold value. Given the complexity of the likelihood function for $SETAR$ models, the exact maximum likelihood is not defined, so the pooled AIC has been derived from the conditional maximum likelihood, i.e.

$$LogL(y \mid \theta) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{j=1}^{k}(T_j \log \sigma_{\varepsilon_j}^2 + \frac{S(\phi_j, \gamma_{j-1}, \gamma_j)}{\sigma_{\varepsilon_j}^2}),$$

being p_j and $\hat{\sigma}_{\varepsilon j}^2$ respectively the lag length and the residual variances for the $j-th$ generic regime, whereas T, ϕ_j , d, γ have already been defined (Eqn. 10) and $S(\phi_j, \gamma_{j-1}, \gamma_j) = \sum_{\gamma_{j-1} \leq y_{t-d} < \gamma_j} \varepsilon_{jt}^2,\ j = 1, ..., l$.

For a generic model, realization of a l-regimes $SETAR$ process, pooled AIC is defined as follows:

$$AIC(p_1, ..., p_l) = \sum_j \left\{ T_j \ln \hat{\sigma}^2_{\varepsilon j} + 2(p_j + 1) \right\}, \quad j = 1, .., l. \tag{12}$$

Being the sum of the AIC's computed for each of the l regimes, this compound statistic is able to balance fitness and complexity by taking into account the peculiar dynamic structures inherent to each state of the process under investigation.

2.3 Model selection procedure

MAICE selection strategy (for a thorough account of this approach, the reader is referred to [26]) is a procedure aimed at extracting, among the candidate models in the competition set $\left\{ \boldsymbol{\mathcal{M}};\ \ \mathfrak{M}_1, \ldots, \mathfrak{M}_m \right\}$, a model structure, $\mathcal{T}_0$ satisfying:

$$\boldsymbol{\mathcal{M}} \supset \mathcal{M}_0 = \arg\min_{\boldsymbol{\Xi}} AIC(\boldsymbol{\mathcal{M}}), \tag{13}$$

where $\boldsymbol{\Xi}$ is as defined in Eqn. 8. MAICE procedure requires the a priori and arbitrary definition of appropriate upper bound for each of the models' meta–parameters. Particularization of (13) for our operative ...framework is straightforward and leads to the definition of their respective competition sets, i.e. $\left\{ {}^{\boldsymbol{a}}\boldsymbol{\mathcal{M}} \right\}$ and $\left\{ {}^{\boldsymbol{b}}\boldsymbol{\mathcal{M}} \right\}$, and the related MAICE procedures, that is ${}^{\boldsymbol{a}}\boldsymbol{\mathcal{M}} \supset^a \mathcal{M}_0 = \arg\min_{{}^a\boldsymbol{\Xi}} AIC({}^{\boldsymbol{a}}\boldsymbol{\mathcal{M}})$ and ${}^{\boldsymbol{b}}\boldsymbol{\mathcal{M}} \supset^b \mathcal{M}_0 = \arg\min_{{}^b\boldsymbol{\Xi}} AIC({}^{\boldsymbol{b}}\boldsymbol{\mathcal{M}})$.

2.4 Optimal Wavelet Filtering and Forecast Combination

Let $\boldsymbol{\gamma} \equiv (\gamma_1, ..., \gamma_\omega)$ be a ω-dimensional vector of competitive wavelet thresholds, we want to find, for each horizon $h \in H$ (Eqn. 2), the H–dimensional vector of "best" threshold parameters, according to a predefined rule. In order to make the explanation simpler, without loss of generality, only one horizon will be considered – kept fixed throughout the paper and equal to $h_0 \equiv OSH$ (One Step Ahead), i.e. $H = h_0$. Also, it will be assumed the vector $\boldsymbol{\gamma}$ to be restricted to only two competitive thresholds parameters and equal for the two models a and b, i.e. $\boldsymbol{\gamma} \equiv \boldsymbol{\delta} \equiv (\gamma_1 = \delta_1, \gamma_2 = \delta_2)$.

MAICE procedure is carried out on the competition sets ${}^{\boldsymbol{a}}\boldsymbol{\mathcal{M}}$ and ${}^{\boldsymbol{b}}\boldsymbol{\mathcal{M}}$, after the model fitting procedure has been conducted on the two filtered versions of the original time series – i.e. $(x_t|\gamma_1 \subset \gamma)$ and $(x_t|\gamma_2 \subset \gamma)$ – in symbols: $MAICE({}^{\boldsymbol{I}}\boldsymbol{\mathcal{M}}(x_t|\gamma_c \subset \gamma));\quad c = 1, 2$, so that the prediction residual variances can be estimated and stored in a matrix, say ${}^I\boldsymbol{\Sigma}$, i.e.

$$\boldsymbol{\Sigma} = \begin{bmatrix} {}^a\hat{\sigma}^2_{\varepsilon}(\gamma_1) & {}^b\hat{\sigma}^2_{\varepsilon}(\gamma_1) \\ {}^a\hat{\sigma}^2_{\varepsilon}(\gamma_2) & {}^b\hat{\sigma}^2_{\varepsilon}(\gamma_2) \end{bmatrix},$$

whose generic entry ${}^{I}\hat{\sigma}^2_{\varepsilon}(x_t|\gamma_c)$ is the conditional residual variance of prediction generated by the selected models of each type, a and b. Column–wise minimization of $\boldsymbol{\Sigma}$, leads to the determination of the optimal thresholds – i.e. ${}^{I}\gamma^* = \min_{\gamma\in\gamma} \hat{\sigma}^2_{\varepsilon}(\boldsymbol{I}_{MAICE})$ – as well as the prediction residual variances.

At this point, the procedure proposed by Bates and Granger is applied and the optimal pair of weights are found – i.e. ω and $(1-\omega)$ – which will be associated to the two prediction sets generated by the two models, for the particular horizon(s) chosen (h_0, in our case). To do so, being ${}^{(I)}\sigma^2_{\varepsilon}$ the objective function, the variance of the combination of the two prediction sets is given by

$$ {}^{(I)}\sigma^2_{\varepsilon} = \ \omega^2 \ {}^{a}\sigma^2_{\varepsilon} + (1-\omega)^2 \ {}^{b}\sigma^2_{\varepsilon} + 2\rho\omega \ {}^{a}\sigma_{\varepsilon} \ (1-\omega){}^{b}\sigma_{\varepsilon}, \tag{14} $$

with $\rho = corr({}^{a}\varepsilon_t \ ; \ {}^{b}\varepsilon_t)$; $corr$ = correlation coefficient. Finally, the minimization of the overall variance ${}^{(I)}\sigma^2_{\varepsilon}$ – achieved by differenciating Eqn. 14 w.r.t. ω and equating the derivative to zero – leads to the determination of the optimal weight ω^*, i.e.

$$ \omega^* = \frac{{}^{b}\sigma^2_{\varepsilon} - \rho \ {}^{a}\sigma^{b}_{\varepsilon}\sigma_{\varepsilon}}{{}^{a}\sigma^2_{\varepsilon} + \ {}^{b}\sigma^2_{\varepsilon} - 2\rho \ {}^{a}\sigma^{b}_{\varepsilon}\sigma_{\varepsilon}} $$

The choice of ω^* can be performed only provided that a certain number of predictions are evaluated, therefore a portion of the available data set (in the sequel referred to as the validation set) must be devoted to this task.

2.5 The algorithm

FOR–WARD procedure is now detailed in a step–by–step fashion. For an easier readability, the algorithm has been broken down into three parts: the first accounts for the a priori choices and the standard procedure – based on the simple combination of the predictions – whereas the second focuses on how the optimal weights ${}^{a}w^*$ and ${}^{b}w^*$ are obtained. The application of the method in real-life and for simulation purposes is illustrated in the third part.

Pre–specification and standard procedure

1. Let $\{x\}_1^T$ (Eqn. 1–2) be the time series of interest, on which no particular assumptions are made. The data are split into three disjoint segments: training set $\{x^{tr}\}_t$; $t = 1, 2, ..., T-(S+V)$, validation set, $\{x^{va}\}_t$ $t = T-(S+V)+1, ..., (T-V)$ and test set, $\{x^{ts}\}_t$ $t = T-V+1, ..., T$), V and S being respectively the length of the validation and the test set;
2. the competition set $\boldsymbol{\mathcal{M}}$, formed by all the arbitrary chosen models (m's), i.e. $\left\{\boldsymbol{\mathcal{M}} \equiv \ {}^{\boldsymbol{a}}\boldsymbol{M} \cup {}^{\boldsymbol{b}}\boldsymbol{M};\right\}$, with $\left\{{}^{I}\boldsymbol{M} \equiv ({}^{a}m_1, {}^{a}m_2, \ldots, {}^{a}m_{M_a}), \quad \{I=a\}\right\}$ and $\left\{{}^{I}\boldsymbol{M} \equiv ({}^{b}m_1, {}^{b}m_2, \ldots, {}^{b}m_{M_b}), \quad \{I=b\}\right\}$, is built;

3. two vectors – i.e. $\{\boldsymbol{\gamma}_\omega;\ \omega = 1, 2, \ldots, \Omega\}$ and $\{\boldsymbol{\delta}_\omega;\ \omega = 1, 2, \ldots, \Omega\}$ – containing the tentative thresholds, respectively for the model a and the model b, are built. Without loss of generality, they are assumed to have the same dimension Ω;
4. the vector $h \equiv (1, 2, ..., H)'$ containing the (arbitrary) forecast horizons is defined;
5. the set ${}^{I}S(\boldsymbol{\Xi});\quad I = (a, b)$ containing reasonable, arbitrarily chosen, boundaries for the model structures, is built;
6. an appropriate waveform is chosen;
7. the models in $\{\boldsymbol{\mathcal{M}}\}$ are iteratively fitted, up to ${}^{I}S(\boldsymbol{\Xi})$ and $\forall h \in H$, on the training set and on its extensions $\{x\}_1^{T-(S+V-1-h)}, \ldots, \{x\}_1^{T-(S-h)}$, which are obtained by adding one at a time the observations belonging to the validation set to $\{x^{tr}\}$;
8. the AIC is extracted – according to Equations 9 and 12, $\forall\ {}^{I}m \in \boldsymbol{{}^{I}M}$, and stored in the two H–dimensional vectors $\boldsymbol{{}^{I}\varrho} = AIC(\boldsymbol{{}^{I}M})$;
9. MAICE procedure is applied to $\boldsymbol{{}^{I}\varrho}$, i.e. $MAICE(\boldsymbol{{}^{a}\varrho})$ and $MAICE(\boldsymbol{{}^{b}\varrho})$, so that the set $\left\{\boldsymbol{\mathcal{M}^*}\right\}$ containing the winner models generated by the model a, say $\boldsymbol{{}^{a}M}^*$ and model b, say $\boldsymbol{{}^{b}M}^*$, is built, i.e.: $\left\{\boldsymbol{\mathcal{M}^*} \equiv \boldsymbol{{}^{a}M}^* \cup \boldsymbol{{}^{b}M}^*\right\} \subset \boldsymbol{\mathcal{M}}$;
10. $\{\boldsymbol{\mathcal{M}^*}\}$ is employed to generate $h = 1, 2, \ldots, H$–step ahead predictions, which are stored in two matrices, i.e. $\left\{{}^{I}\boldsymbol{P};\quad I = (a, b)\right\}$, both of dimension $V \times H$, i.e.

$$\boldsymbol{{}^{I}\hat{P}_h} \equiv \begin{bmatrix} \hat{x}^{h=1}_{T-(S+V-1)+1} & \cdots & \cdots & \cdots\ \cdots & \cdots \\ \hat{x}^{h=1}_{T-(S+V-1)+2} & \hat{x}^{h=2}_{T-(S+V-1)+2} & \cdots & \cdots\ \cdots & \cdots \\ \hat{x}^{h=1}_{T-(S+V-1)+3} & \hat{x}^{h=2}_{T-(S+V-1)+3} & \hat{x}^{h=3}_{T-(S+V-1)+3} & \cdots\ \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots\ \cdots & \cdots \\ \hat{x}^{h=1}_{T-(S+V-1)+H} & \hat{x}^{h=2}_{T-(S+V-1)+H} & \hat{x}^{h=3}_{T-(S+V-1)+H} & \cdots\ \cdots & \hat{x}^{h=H}_{T-(S+V-1)+H} \\ \\ \cdots & \cdots & \cdots & \cdots\ \cdots & \cdots \\ \hat{x}^{h=1}_{T-S} & \hat{x}^{h=2}_{T-S} & \hat{x}^{h=3}_{T-S} & \cdots\ \cdots & \hat{x}^{h=H}_{T-S} \end{bmatrix};$$

11. $\forall h \in H$, the residual variances of the validation set, i.e. $\boldsymbol{{}^{I}\Pi_H} \equiv E(\boldsymbol{x}^{va} - \boldsymbol{{}^{I}P})^2$ – being $\boldsymbol{x}^{va}$ the $V \times H$ matrix of the observed and shifted data (equal to $\boldsymbol{{}^{I}\hat{P}_h}$ but with the observed data instead of their estimates) and E the expectation operator – are stored in the H–dimensional vector $\boldsymbol{{}^{I}\Pi_H}$, whose generic element is given by

$${}^{I}\hat{\sigma}^2_\varepsilon(h) = \frac{1}{V}\sum_{v\in V}(x_{v+h} - {}^{I}\hat{x}_{v+h})^2 = \frac{1}{V}\sum_{v\in V}({}^{I}\varepsilon_{v+h})^2;\ \forall h = 1, \ldots, H;\ I = (a, b); \tag{15}$$

12. validation set optimal weights ${}^{a}\omega_h$ and ${}^{b}\omega_h$, for each forecast horizon h, are computed according to

$${}^{a}\omega_h = \frac{{}^{b}\hat{\sigma}^2_\varepsilon(h) - \rho_h[{}^{a}\hat{\sigma}_\varepsilon(h)\,{}^{b}\hat{\sigma}_\varepsilon(h)]}{{}^{b}\hat{\sigma}^2_\varepsilon(h) + {}^{a}\hat{\sigma}^2_\varepsilon(h) - 2\rho_h[{}^{a}\hat{\sigma}_\varepsilon(h)\,{}^{b}\hat{\sigma}_\varepsilon(h)]};\quad {}^{b}\omega_h = 1 - {}^{a}\omega_h,\quad \forall h \in H, \tag{16}$$

where ρ_h is the correlation coefficient between the h–step ahead errors generated by the two set of predictions;

13. the pair $({}^a\omega_h, {}^b\omega_h)$ is applied to the h–step ahead, out-of-sample predictions provided by the models a and b, i.e.

$$\hat{x}_{T-V+h} \;=\; {}^a\omega_h \hat{x}_{T-V+h} \;+\; {}^b\omega_h \hat{x}_{T-V+h} \quad \forall h \in H; \tag{17}$$

14. the remaining test data are sequentially drawn, one at a time, from $\{x^{ts}\}_t \quad t = T-(V-h-1), ..., T)$ and used to iteratively recompute the weights. In practice, steps 7 – 12 are repeated $S-1$ times. The related predictions are then generated according to Eqn. 17;
15. a quadratic loss function, i.e.

$$\mathcal{L}(\omega) = E(x_{T-V+h} - \hat{x}_{T-V+h})^2, \tag{18}$$

is computed.

Training and Validation

15. Two tentative threshold parameters, say $\gamma_0 \in \gamma$ and $\delta_0 \in \delta$, are selected;
16. the training set is filtered according to γ_0 and δ_0, i.e. $x_t^{tr}(\gamma_0)$ and $x_t^{tr}(\delta_0)$;
17. conditioning to the first pair of thresholds, the vector of the H residual variances is obtained applying Eqn. 15, i.e.:

$${}^I\hat{\boldsymbol{\sigma}}^{\boldsymbol{2}}_{\boldsymbol{\varepsilon}}(h_j, \gamma_0, \delta_0) \;\equiv [\; {}^I\hat{\sigma^2_\varepsilon}(1, \gamma_0, \delta_0),\; {}^I\hat{\sigma^2_\varepsilon}(2, \gamma_0, \delta_0), \ldots,\; {}^I\hat{\sigma^2_\varepsilon}(H, \gamma_0, \delta_0)]; \tag{19}$$

18. steps 7 – 11 are repeated for the remaining $\Omega - 1$ thresholds so that the $\Omega \times H$ matrices ${}^a\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}}(h, \gamma_\omega)$ and ${}^b\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}}(h, \delta_\omega)$, containing the residual variances for each of the thresholds and horizons considered, are generated, i.e.:

$${}^{\boldsymbol{a}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}} \equiv \begin{bmatrix} {}^a\hat{\sigma}^2_\varepsilon(1,\gamma_1) & {}^a\hat{\sigma}^2_\varepsilon(2,\gamma_1) & \ldots\ {}^a\hat{\sigma}^2_\varepsilon(H,\gamma_1) \\ {}^a\hat{\sigma}^2_\varepsilon(1,\gamma_2) & {}^a\hat{\sigma}^2_\varepsilon(2,\gamma_2) & \ldots\ {}^a\hat{\sigma}^2_\varepsilon(H,\gamma_2) \\ \ldots & \ldots & \ldots \quad \ldots \\ {}^a\hat{\sigma}^2_\varepsilon(1,\gamma_\Omega) & {}^a\hat{\sigma}^2_\varepsilon(2,\gamma_\Omega) & \ldots\ {}^a\hat{\sigma}^2_\varepsilon(H,\gamma_\Omega,) \end{bmatrix}$$

and

$${}^{\boldsymbol{b}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}} \equiv \begin{bmatrix} {}^b\hat{\sigma}^2_\varepsilon(1,\delta_1) & {}^b\hat{\sigma}^2_\varepsilon(2,\delta_1) & \ldots\ {}^b\hat{\sigma}^2_\varepsilon(H,\delta_1) \\ {}^b\hat{\sigma}^2_\varepsilon(1,\delta_2) & {}^b\hat{\sigma}^2_\varepsilon(2,\delta_2) & \ldots\ {}^b\hat{\sigma}^2_\varepsilon(H,\delta_2) \\ \ldots & \ldots & \ldots \quad \ldots \\ {}^b\hat{\sigma}^2_\varepsilon(1,\delta_\Omega) & {}^b\hat{\sigma}^2_\varepsilon(2,\delta_\Omega) & \ldots\ {}^b\hat{\sigma}^2_\varepsilon(H,\delta_\Omega,) \end{bmatrix};$$

19. the H–dimensional vectors $\boldsymbol{\gamma}^*_{\boldsymbol{H}} \equiv (\gamma^*_1, \ldots, \gamma^*_H)'$ and $\boldsymbol{\delta}^*_{\boldsymbol{H}} \equiv (\delta^*_1, \ldots, \delta^*_H)'$ of optimal thresholds – i.e. conditioned to each horizon h – is generated by applying column-wise the minimization operation on ${}^{\boldsymbol{I}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}}$, i.e. $(\gamma^* \mid h_j) = \arg\min_{\gamma}({}^I\boldsymbol{\Sigma}_{(\cdot\ ;\ h)})$; and $(\delta^* \mid h_j) = \arg\min_{\delta}({}^I\boldsymbol{\Sigma}_{(\cdot\ ;\ h)})$, for $I = a, b$ respectively;
20. the optimal pair ω^*_h of weights, for each horizon h is found through

$$ {}^a\omega^*_h = \frac{{}^b\hat{\sigma}^2_{\varepsilon}(\delta^*;h) - \rho_h[{}^a\hat{\sigma}_{\varepsilon}(\gamma^*;h){}^b\hat{\sigma}_{\varepsilon}(\delta^*;h)]}{{}^b\hat{\sigma}^2_{\varepsilon}(\delta^*;h) + {}^a\hat{\sigma}^2_{\varepsilon}(\gamma^*;h) - 2\rho_h[{}^a\hat{\sigma}_{\varepsilon}(\gamma^*;h){}^b\hat{\sigma}_{\varepsilon}(\delta^*;h)]}; \quad {}^b\omega^*_h = 1 - {}^a\omega^*_h, \ \forall h \in H \tag{20} $$

where ρ_h is as already defined (Eqn. 16).

Performances' assessment and Real-life employment

21. The pair $({}^a\omega^*_h, {}^b\omega^*_h)$ is applied to the h–step ahead, out-of-sample predictions provided by the models a and b, i.e.

$$ \hat{x}^*_{T-V+h} \ = \ {}^a\omega^*_h\hat{x}_{T-V+h}(\gamma^*_h) \ + \ {}^b\omega^*_h\hat{x}_{T-V+h}(\delta^*_h) \quad \forall h \in H; \tag{21} $$

22. the remaining test data are sequentially drawn one at a time from $\{x^{ts}\}_t \quad t = T-(V-h-1), ..., T)$, and used to iteratively recompute the weights. In practice, steps 15 – 21 are repeated $S-1$ times. The related predictions are then generated according to Eqn. 21;
23. a quadratic loss function, i.e. $\mathcal{L}^*(\omega) = E(x_{T-V+h} - \hat{x}^*_{T-V+h})^2$, is computed;
24. under $\mathcal{L} > \mathcal{L}^*$ (Eqn. 18), FOR–WARD procedure is chosen and real-life predictions are generated according to

$$ \hat{x}_{T+h} \ = \ {}^a\omega^*_h\hat{x}_{T+h}(\gamma^*_h) \ + \ {}^b\omega^*_h\hat{x}_{T+h}(\delta^*_h) \quad \forall h \in H. $$

3 Empirical Analysis.

In this section, the outcomes of an empirical study, based on four macroeconomic time series, are presented. They are: Civilian Labor Force Participation Rate (Percent, Monthly, Not Seasonally Adjusted), ISM Manufacturing: New Orders Index© (Monthly, Seasonally Adjusted), Civilian Unemployment Rate (Percent, Monthly, Not Seasonally Adjusted), Effective Federal Funds Rate (Percent, Monthly, Not Seasonally Adjusted). These series, in the sequel denoted respectively as $TS1$, $TS2$, $TS3$, $TS4$, are graphically portrayed in Figure 1 and detailed in Table 1, whereas their empirical autocorrelation functions ($EACF$) are displayed in Figure (2). They have been selected for the substantial differences shown in many regards such as type of phenomenon measured, inherent characteristics (e.g. time span, probabilistic structure, seasonality and frequency components) and degree of smoothness. The latter is high in the case of TS4, whereas in TS3 the high frequency components seem to superimpose the signal and to generate a pronounced overall roughness as a result.

The analysis of the dynamics suggests, for TS1-3-4, a non stationary pattern (especially TS4), with some evidences of multiregime behavior, whereas TS2 show a mean reverting behavior and a tendency towards omoschedasicity. Finally, Figure 2 emphasizes rapidly decaying memory structures for TS2 and, to a less extent, TS3 whereas the remaining series exhibit long term linear memory behavior. In order to quantitatively measure the quality of the predictions generated, the following three metrics – computed on the test set $\{x^{tst}\}_{t\in\mathbb{Z}^+}$) – will be employed: $RMSE^{(h)} = \sqrt{\frac{1}{S}\sum|x^{tst} - \hat{x}^{tst}|^2}$, $MPE^{(h)} = 100\frac{1}{S}\sum[\frac{x^{tst}-\hat{x}^{tst}}{x^{tst}}]$, $MAPE^{(h)} = 100\frac{1}{S}\sum|\frac{x^{tst}-\hat{x}^{tst}}{x^{tst}}|$, S, as above specified, being the length of x^{tst} and h_i; $\quad i = 1, .., 4$.

Regarding the waveform, its choice is in general system–dependent and hard to automatize. In the case of the considered time series, an analysis of their characteristics, along with an approach of the type trial-and- error, has lead to the choice of a nearly symmetric wavelet of the type Symmlet S8, with 8 vanishing moments. A thorough investigation of this popular type of wavelet may be found in many sources (e.g. [27]); [28]). Finally, the threshold scheme adopted, formalized in Eqn. 6, has been discussed in [29] and [30].

Name	Variable	Source	Period	Span
$TS1$	Civilian Labor Force Participation Rate	Federal Reserve Bank of St. Louis (Economic Research Division)	Jan. 1980 – Jan. 2015	432
$TS2$	ISM Manufacturing: New Orders Index	Institute for Supply Management (Manufacturing ISM Report on Business)	Jan. 1985 – Jan. 2015	372
$TS3$	Civilian Unemployment Rate	US Bureau of Labor Statistics (Household Survey)	Jan. 1990 – Dec. 2015	312
$TS4$	Effective Federal Funds Rate	Federal Reserve Bank of St. Louis	Jan. 1990 – Dec. 2015	312

Table 1: The time series employed in the empirical analysis: main details

3.1 Results

The results of the empirical experiment are now reported, along the arbitrary choices made on the model hyperparameters and the employed grid sets. To maintain the computational time at acceptable levels, $\boldsymbol{\gamma} = \boldsymbol{\delta}$ sets' cardinality has been kept small and fixed to 6 for both the models a and b, i.e. $\boldsymbol{\gamma} \equiv (\gamma_1 = .1, \gamma_2 = .4, \gamma_3 = .6, \gamma_4 = .8, \gamma_5 = 1.1, \gamma_6 = 1.5)$. A two–regimes SETAR model has been considered, with maximum order set to 6 for both the regimes and with

delay parameter ranging from 0 to 2. The SARIMA model has its maximum dimensionality set to 3 for the regular AR and MA parameters and to 2 for the seasonal part. Maximum order of integration is d=2 (for the regular difference) and D=1 (for the seasonal difference). The segmentation of the data sets (training, validation and test) is as reported in Table 2. Finally, as a benchmark, the unfiltered version of the method – i.e. generating the compounded predictions using the original time series only with no filtering – has been employed.

The good performance of FOR-WARD procedure can be noticed by inspecting Tables 3 and 4. In particular, the predictions delivered at the shortest horizons (h=1,2) can be considered interesting: looking at TS1-2, the RMSE values are equal to .103 and .550 for the filtered time series versus .205 and 1.007 respectively. The values for h=2 are .199 and .931 versus .277 and 1.322. The higher the horizon the less impressive seem to be FOR-WARD's performances. However, worth of mentioning seems the fact that the values for the MAPE and the MPE recorded at horizon h=3 are still consistently better than those yielded by the unfiltered version of the model. TS4 is the case where FOR–WARD performances are basically in line with those of its competitor: they are in general very close to each other and, in some cases, even slightly worse. This can be explained by the intrinsic characteristic of the time series itself, which seems to be quite smooth and, as such, not susceptible to benefit from the filtering capabilities delivered by the proposed procedure. This fact is confirmed by the small values selected for the thresholds at all the considered horizons. Conversely, as expected, the other thresholds show bigger threshold values, see for example TS3, whose roughness has been above pointed out.

Size / Set	$TS1$	$TS2$	$TS3$	$TS4$
Training	372	312	264	264
Validation	36	36	24	24
Test	24	24	24	24

Table 2: Size of each subset the original time series have been broken into

	Horizon	$^a\gamma^*$	$^b\gamma^*$	RMSE	MPE	MAPE		Horizon	$^a\gamma^*$	$^b\gamma^*$	RMSE	MPE	MAPE
	1	.3	.3	.103	.062	.129		1	.2	.2	.550	.300	.842
TS1	2	.2	.2	.199	- .112	.233	*TS2*	2	.2	.1	.931	.308	1.38
	3	.3	.3	.318	.267	.440		3	.3	.3	2.313	1.255	3.368
	4	.5	.3	.461	.447	.639		4	.8	.8	2.853	2.842	4.140
	1	.5	.5	0.189	-0.453	2.619		1	.1	.1	.017	.695	14.133
TS3	2	.5	.5	0.295	-3.664	4.791	*TS4*	2	.1	.1	.0370	7.499	25.025
	3	.8	.8	.403	- .729	5.712		3	.1	.1	.051	9.989	33.983
	4	.8	.8	0.971	11.327	12.375		4	.1	.1	.062	-7.09	55.564

Table 3: Goodness of fit statistics computed on the test set for the four time series considered – Filtered case

	Horizon	RMSE	MPE	MAPE		Horizon	RMSE	MPE	MAPE
	1	.205	.147	.282		1	1.007	1.522	1.576
TS1	2	.277	.095	.367	*TS2*	2	1.322	.690	1.949
	3	.414	.322	.570		3	2.972	1.792	4.301
	4	.560	.421	.641		4	2.457	1.420	3.571
	1	0.380	-1.550	5.435		1	.017	.792	14.093
TS3	2	0.432	-4.796	6.908	*TS4*	2	.032	7.907	22.138
	3	0.506	1.888	6.517		3	.043	13.445	27.372
	4	1.384	11.442	16.980		4	.062	-7.178	55.624

Table 4: Goodness of fit statistics computed on the test set for the four time series considered – Unfiltered case

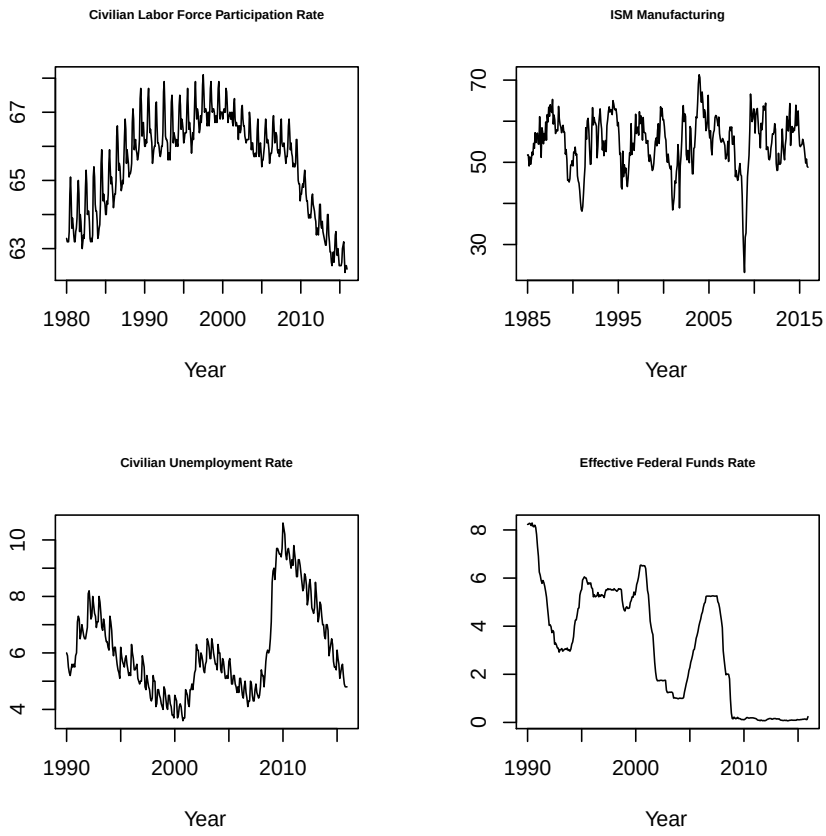

Fig. 1: Graphical representation of the time series used in the empirical experiment

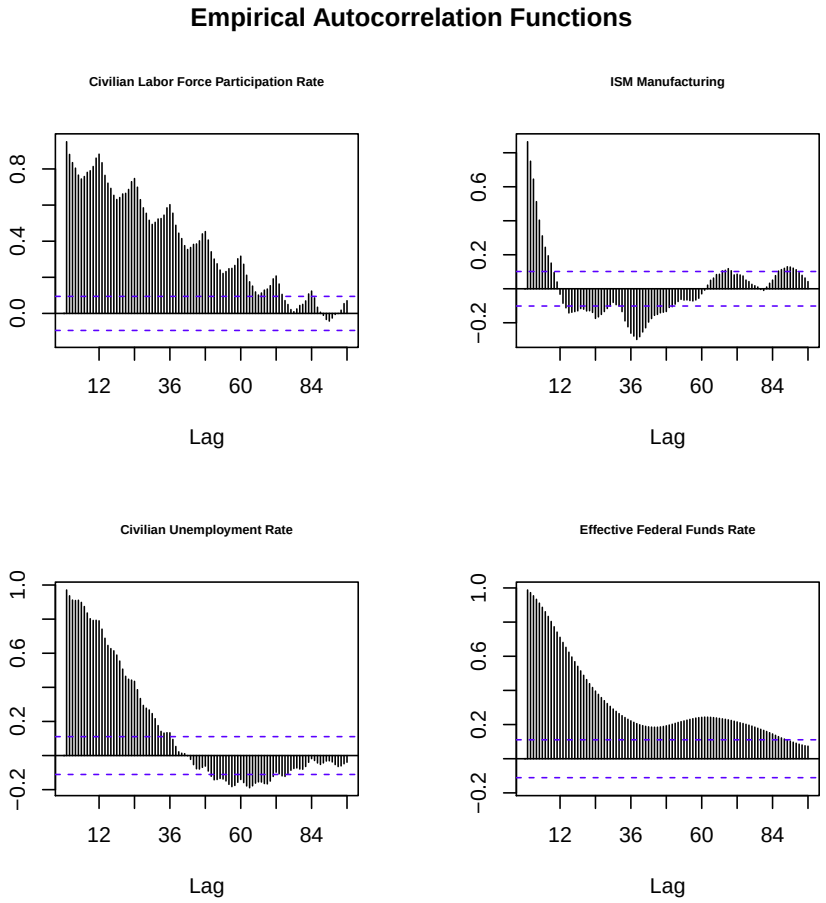

Fig. 2: Graphical representation of the time series used in the empirical experiment

Acknowledgment Part of the elaborations have been performed using the computing resource "Pythagoras", of the Mathematical Department of the University of California San Diego. Special thanks go to Dr. Andrea Stanco (Istat), for his help in designing and implementing the program used for the simulation process and for the valuable suggestions provided.

References

1. Bates JM, Granger CW. The combination of forecasts. Or. 1969;p. 451–468.
2. Box GEP, Jenkins GM. Time Series Analysis, Forecasting, and Control. San Francisco: Holden-Day; 1976.
3. Tong H. Some comments on the Canadian lynx data. Journal of the Royal Statistical Society Series A (General). 1977;p. 432–436.
4. Tong H, Lim KS. Threshold autoregression, limit cycles and cyclical data. Journal of the Royal Statistical Society Series B (Methodological). 1980;p. 245–292.
5. Soares LJ, Medeiros MC. Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data. International Journal of Forecasting. 2008;24(4):630–644.
6. Chakhchoukh Y, Panciatici P, Bondon P. Robust estimation of SARIMA models: Application to short-term load forecasting. In: 2009 IEEE/SP 15th Workshop on Statistical Signal Processing. IEEE; 2009. p. 77–80.
7. Sigauke C, Chikobvu D. Prediction of daily peak electricity demand in South Africa using volatility forecasting models. Energy Economics. 2011;33(5):882–888.
8. Brida JG, Risso WA. Research note: Tourism demand forecasting with SARIMA models–the case of South Tyrol. Tourism Economics. 2011;17(1):209–221.
9. Saz G. The efficacy of SARIMA models for forecasting inflation rates in developing countries: the case for Turkey. International Research Journal of Finance and Economics. 2011;62:111–142.
10. Boero G, Marrocu E. The performance of SETAR models: a regime conditional evaluation of point, interval and density forecasts. International Journal of Forecasting. 2004;20(2):305–320.
11. Clements MP, Smith J. Evaluating forecasts from SETAR models of exchange rates. Journal of International Money and Finance. 2001;20(1):133–148.
12. Kräger H, Kugler P. Non-linearities in foreign exchange markets: a different perspective. Journal of International Money and Finance. 1993;12(2):195–208.
13. Juvenal L, Taylor MP. Threshold adjustment of deviations from the law of one price. Studies in Nonlinear Dynamics & Econometrics. 2008;12(3).
14. Potter SM. A nonlinear approach to US GNP. Journal of applied econometrics. 1995;10(2):109–125.
15. Dufrénot G, Guegan D, Peguin-Feissolle A. Long-memory dynamics in a SETAR model–applications to stock markets. Journal of International Financial Markets, Institutions and Money. 2005;15(5):391–406.
16. Jamaleh A. " A Threshold Model for Italian Stock Market Volatility. Potential of These Models: Comparing Forecasting Performances and Evaluation of Portfolio-Risk. Working paper, Revista Politica Economia; 2001.
17. Stoica P, Selen Y. Model-order selection: a review of information criterion rules. IEEE Signal Processing Magazine. 2004;21(4):36–47.
18. Choi B. ARMA model identification. Springer Science & Business Media; 2012.
19. Akaike H. A new look at the statistical model identification. IEEE transactions on automatic control. 1974;19(6):716–723.
20. Hannan EJ, Quinn BG. The determination of the order of an autoregression. Journal of the Royal Statistical Society Series B (Methodological). 1979;p. 190–195.
21. Ozaki T. On the order determination of ARIMA models. Applied Statistics. 1977;p. 290–301.

22. Akaike H. Time series analysis and control through parametric models. Applied Time Series Analysis. 1978;p. 1–23.
23. Stone CJ. Local asymptotic admissibility of a generalization of Akaike's model selection rule. Annals of the Institute of Statistical Mathematics. 1982;34(1):123–133.
24. Di Narzo F. Nonlinear autoregressive time series models in R using tsDyn version 0.7. R documentation. 2008;.
25. Fenga L, Politis DN. Bootstrap order selection for SETAR models. Journal of Statistical Computation and Simulation. 2015;85(2):235–250.
26. Akaike H. Prediction and entropy. In: Selected Papers of Hirotugu Akaike. Springer; 1985. p. 387–410.
27. Baleanu D, Aydin H. Advances in wavelet theory and their applications in engineering, physics and technology. InTech; 2012.
28. Härdle W, Kerkyacharian G, Picard D, Tsybakov A. Compactly supported wavelets. In: Wavelets, Approximation, and Statistical Applications. Springer; 1998. p. 59–69.
29. Härdle W, Kerkyacharian G, Picard D, Tsybakov A. Wavelets, approximation, and statistical applications. vol. 129. Springer Science & Business Media; 2012.
30. Joshi S, Mukane DS. Comparative Analysis of Thresholding Techniques using Discrete Wavelet Transform. International Journal of Electronics Communication and Computer Engineering. 2014;5(4).