

Smoothing parameter estimation for first order discrete time infinite impulse response filters

Abstract

Discrete time Infinite Impulse Response low-pass filters are widely used in many fields such as engineering, physics and economics. Once applied to a given time series, they have the ability to pass low frequencies and attenuate high frequencies. As a result, the data are expected to be less noisy. A properly filtered signal, is generally more informative with positive repercussions involving qualitative aspects – e.g. visual inspection and interpretation – as well as quantitative ones, such as its digital processing and mathematical modelling. In order to effectively disentangle signal and noise, the filter smoothing constant, which controls the degree of smoothness in First Order Discrete Time Infinite Impulse Response Filters, has to be carefully selected. The proposed method conditions the estimation of the smoothing parameter to a modified version of the information criterion of the type Hannan - Quinn which in turns is built using the Estimated Log Likelihood Function of a model of the class SARIMA (Seasonal Auto Regressive Moving Average). Theoretical evidences as well as an empirical study conducted on a particularly noisy time series will be presented.

Keywords: big data, discrete time infinite impulse response filters, denoising, hannan quinn information criterion, sarima models, time series

Volume 7 Issue 5 - 2018

Livio Fenga

Italian National Institute of Statistics, Italy

Correspondence: Livio Fenga, Italian National Institute of Statistics, Italy, Tel +3906 4673 3222, Email lfenga@istat.it

Received: August 27, 2018 | **Published:** October 22, 2018

Introduction

Among the many denoising methods and techniques successfully employed for univariate time series – e.g. based on regression,¹ Kalman filter,^{2,3} decomposition,⁴ wavelet^{5,6} and non-linear method⁷ – those based on algorithms of the type Infinite Impulse Response (IIR) exponential filters have been massively used, given their satisfactory performances (see, for example,⁸ and, more recently⁹). Such methods are useful for their ability to maximize the amount of relevant information that can be extracted from “real life” time series. In fact, regardless the scientific field time dependent data are collected for (e.g. engineering, economics, physics, environmental), they can never be error-free. In spite of all of the efforts and precautions one might take in order to provide clean data – e.g. robust data acquisition methods, reliable routine checks, sophisticated procedures for error correction, fail safe data storage and data communication lines – reality is way too complex for such procedures to be completely reliable.

Noise, in fact, is simply an ubiquitous entity able to affect virtually all the stages a given analysis of a time series can be broken into, showing uncountable expressions that can be only partially controlled, never fully removed nor exactly pinpointed. Many are the fields where noise reduction methods are employed: data mining, satellite data, radio communications, radar, sonar and automatic speech recognition, are just a few of them. Often, the treatment of noisy data is critical, as in the case of bio-medical signals, tracking systems for defense purposes or economics, where the trend or the seasonal patterns of vital economic variables can be obscured or distorted by noise components. However, the type of time series likely to be easily affected by noise – as well as by sudden or unpredictable high frequency variations – can be found in the economic and social fields. A telling example, is the great deal of data generated by web-based services (e.g. Google or Twitter), which are commonly and publicly available, in many instances free of charge. However, statistical estimation procedures based on this type of data can embody many sources of uncertainty and instability. They are related, for example, to

technical (computer alphabetization) or psychological (e.g. emotional driven spikes and other form of data perturbations) reasons. Even when data are collected and validated by Official Organisms (e.g. national and super-national statistical offices or central banks), strong psychological components can play a significant role in determining erratic and/or noisy behaviors. This is the case, for instance, of the Economic Sentiment Indicators – provided by many National Institute of Statistics – which are purposely designed to capture the amount of optimism (pessimism) towards the future behavior of a set of economic variables. Being these data able to reflect, at least to some extent, the future decision making strategies of a population of reference, they can show dangerous instabilities and irregularities. In other words, when an opinion is in the process to form or change then many analysts are mostly interested in its future developments; however, this is usually the points in time where the data show complex structures, which are usually hard to capture.

The need to conduct more precise model-based investigations under noisy data structures has been the motivating factor of the proposed method. As it will be seen, it delivers interesting performances and is reasonably fast, the latter characteristic making the proposed method a viable option with high dimensional data sets, e.g. of the type big data. It is also risk-free, i.e. when there are no filtering-related benefit – according to a suitable objective function – the method is designed to simply leave the data unfiltered (the λ –smoothing parameter is set to 1). The paper proceeds as follows: in Section 2 the problem is defined along with the statistical model assumed for the underlying data generating mechanism, whereas in Section 3 the employed low-pass filter is illustrated, with a special focus on its calibration issues. The method is explained in Section 4 and the step-by-step implementation of the algorithm is given in paragraph 4.3. Finally, an empirical experiment, conducted on a time series provided by the search engine Google and recovered in the repository called Google Trends (for Italy the address is <https://trends.google.it/trends/?geo=IT>), is presented in section 5.

Statement of the problem

Throughout the paper, the time series (also defined as signal) of interest is intended to be a real-valued, uniformly sampled, sequence of data points of length T , denoted as:

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

Each datum x_t , being observed with an error, say ε_t , is called x_t^o in contrast with the ‘‘pure’’, noise-free, unobservable one x_t^u . Therefore, the formal set up is of the type

$$x_t^o = x_t^u + \varepsilon_t.$$

It is worth stressing that the theoretical framework we are dealing with, does not envision ε_t as being the standard 0-mean constant variance σ^2 idiosyncratic component. In fact, ε_t can also embody all the unpredictable components arising when ‘‘real-life’’ signals come into play and which, in general, are able to generate unwanted phenomena, like sudden outburst of energy (outliers) and noise. As a consequence, at least two detrimental effects are embedded in the signal and can affect its statistical properties: accuracy reduction and generation of heavy noisy components. The former relates to the addition (subtraction) of significant quantities to (from) x_t , so that we have $x^o = kx^u$; $k \neq 1$, while the latter is a subtler, but nonetheless dangerous, disturbing phenomenon. It arises, for example, when in a survey some statistical units are not longer available and must be replaced by ‘‘similar’’ ones. When such a corrupted information set is applied as an input to a mathematical model for any purposes, e.g. simulation or prediction, the related outcomes are likely to be affected.

The data generating process

The observed signal is supposed to be a realization of a statistical model of the class *SARIMA* (Seasonal Auto Regressive Integrated Moving Average), which is a generalization of the *ARIMA* (Auto Regressive Integrated Moving Average) class proposed by George et al.,¹⁰ have been introduced to model complex dynamics i.e. of the type stochastic seasonal. It can be expressed as:^{10,11}

$$\phi_p(B)\Phi_p(B^S)\nabla_S^D\nabla^d X_t = \theta_q(B)\Theta_Q(B^S)\varepsilon_t$$

where B denotes the backward shift operator, d and D denote the non-seasonal and seasonal difference operator respectively and:

$$\nabla^d = 1 - B^d;$$

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p;$$

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q;$$

$$\Phi_p(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \dots - \Phi_p B^{pS};$$

$$\Theta_Q(B^S) = 1 - \Theta_1 B^S - \Theta_2 B^{2S} - \dots - \Theta_Q B^{QS};$$

with ϕ , θ , Φ , Θ , respectively, the non seasonal autoregressive and moving average parameters and the seasonal autoregressive and moving average parameters. Finally ε_t is a 0-mean and finite variance white noise. The model can be estimate when the stationary and invertibility conditions are meet for both the autoregressive and moving average polynomials respectively, that is $\phi_p(B)\Phi_p(B) = 0$ and $\theta_q(B)\Theta_Q(B) = 0$ have their root lying outside the unit circle. Generally, this model is abbreviated as *SARIMA*(p, d, q)(P, D, Q)^S.

When the process is stationary and no seasonal patterns are detected, the model collapses to a pure *ARMA*(p, q) model. The explanation of the *ARIMA* theory is outside the scope of this paper, therefore the interested reader is referred to Makridakis⁸ and Stock.¹²

Under Gaussianity, the *SARIMA* parameter vector

$$G \equiv (\phi_p, \theta_q, \Phi_p, \Theta_Q, \sigma^2) \tag{2}$$

can be estimated via Conditional *MLE* (Maximum Likelihood Estimation) method, which:

In the version employed to run the empirical experiment (Section 5), uses the joint probability density function for $\varepsilon_1, \dots, \varepsilon_T$, i.e.

$$f(\varepsilon_1, \dots, \varepsilon_T | G) = (2\pi\sigma^2)^{-T/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=1}^T \varepsilon_t^2\right\};$$

and

i) Maximizes the function given by

$$\log f(x | \varepsilon_0 = 0; G) = -\frac{T}{2} \log 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^T \varepsilon_t^2,$$

with $[\varepsilon_t; t = 1, \dots, T]$ being recursively estimated using

$$\varepsilon_t = x_t - G\varepsilon_{t-1}. \tag{3}$$

Equation (3) holds by equating the first $(P + Q)$ observations to 0, i.e. $y_0, y_{-1}, \dots, y_{-(P+Q)} = 0$ and setting the first $Q-1$ innovations to 0: $\varepsilon_0, \varepsilon_{-1}, \dots, \varepsilon_{-(Q-1)} = 0$.

The employed filter

By the Central Limit Theorem, by ‘‘properly’’ averaging subsequences of a noisy signal, a better approximation of the ‘‘true’’ ones is obtained as a result.¹³ Consistently, one strategy for noise control relies on a filter of the type discrete time Infinite Impulse Reponse.¹⁴ This class of filters is a powerful tool for handling noise in many fields: in¹⁵ its performances has been compared with many other methods on environmental data (other than simulated ones) whereas its application for standard control-chart procedures is documented in Layth et al.,¹⁶ and Stuart Hunter.¹⁷ Its usefulness for interactive systems and information extraction for complex turbulent flows has been discussed respectively in Casiez¹⁸ and Adrien Cahuzac,¹⁹ whereas the effectiveness of smoothing-driven approaches in economic time series forecasting is documented in the excellent book from Hyndman et al.,²⁰ and in Codrut.²¹

IIR filters envision the level of the stochastic process (1) at a given time t , say ξ_t to be dependent to the present and past observations according to

$$\xi_t = \lambda x_t + \lambda(1-\lambda)x_{t-1} + \lambda(1-\lambda)^2 x_{t-2} + \dots = \lambda \sum_{j=0}^{\infty} (1-\lambda)^j x_{t-j}.$$

Using the lag operator L , such that $L^k x_t = x_{t-k}$, (4) can be re-expressed as

$$\xi_t = \lambda[1 + (1-\lambda)B + (1-\lambda)^2 B^2 + \dots] = [1 - (1-\lambda)B]^{-1} \tag{4}$$

and rewritten as

$$\xi_t = [1 - (1 - \lambda)B]\xi_t = \lambda x_t. \tag{5}$$

From (5) it is evident that ξ_t is a weighted average of x_t (current datum) and the previous level ξ_{t-1} , whose weight sequence is determined by the smoothing parameter λ . Finally, (4) can be expressed, by skipping the lag operator, in a recurrence form as

$$\xi_t = \lambda x_t + (1 - \lambda)\xi_{t-1}; \quad \xi_0 = x_0, \tag{6}$$

or in a error correction form, i.e. determined by the equation $e_t = (x_t - \lambda x_{t-1})$, as follows:

$$\xi_t = \xi_{t-1} + \lambda e_t.$$

The form (6) is the one used in the sequel. Here, the first term represents the contribution added by each and every datum updating our time series whereas the second one accounts for the inertia from the previous observations. Therefore, the amount of noise left in the signal as well as the system lag (the responsiveness of the system output to changes in the input) are both proportional to λ .

It is worth emphasizing that the order of a low-pass filter reflects the amount of strength it uses to attenuate a predetermined frequency band. First order filters of the type (4) reduce the signal amplitude by half for each frequency doubling while higher order filters amplify this effect proportionally with their order. For the purpose pursued here, the order 1 appeared to be appropriate since the filter is:

[(a)]

1. "Only" required to denoise (e.g. no forecasting purposes are pursued);
2. Easier to tune;
3. More interpretable.

However, properly tuning the constant parameter is not a trivial task nor something that – at least in general – can be left to subjective judgment. In fact, a too high λ can leave significant amount of idiosyncratic components in the signal while too low values determine the speed at which the older data are dampened to be too slow (over smoothing).

The smoothing constant and the objective function

As already pointed out, the smoothing parameter (also called smoothing factor or smoothing constant), controls the amount of past information contributing to the formation of the present signal level. Formally, a point in time distant $(t+1) - k = m$ lags influences the level of the signal at time $t+1$ by an amount given by $\xi_{t+1} = \lambda(1 - \lambda)^{t+1-m}$. Even though the proposed method is designed to provide an estimate of λ , it might be interesting to investigate the frequency property of the filter, in order to obtain meaningful information from the selected λ . The related cut-off frequency can be derived^{15,22} by z -transforming (6), i.e.

$$\hat{\xi}(z) = \frac{\lambda}{1 - (1 - \lambda)z^{-1}} \hat{x}(z),$$

which can be regarded as the digital version of the analog filter (6). By defining Δ_t as the sampling period and equating z to the frequency ω , i.e. $z = e^{i\omega\Delta_t}$; $i = \sqrt{-1}$, the power spectra of ξ_t and x_t , respectively called $|\hat{\xi}(\omega)|^2$ and $|\hat{x}(\omega)|^2$ are related to each

other through

$$|\hat{\xi}(\omega)|^2 = \frac{\lambda^2}{|e^{i\omega\Delta_t} - (1 - \lambda)|^2} |\hat{x}(\omega)|^2. \tag{7}$$

By equating the first term in (7) to $\frac{1}{2}$, the cut-off frequency at which the amplitude is reduced by a factor of 2, say \tilde{F} , is found, i.e.

$$\lambda = \frac{1}{3} \sqrt{6(1 - \cos(\tilde{\omega}\Delta_t) + (1 - \cos(\tilde{\omega}\Delta_t))^2 - (1 - \cos(\tilde{\omega}\Delta_t)))} \approx \frac{2\pi\tilde{F}\Delta_t}{\sqrt{3}} \approx 3.62579\tilde{F}\Delta_t,$$

and therefore

$$\lambda = \frac{1}{3} \sqrt{6(1 - \cos(\tilde{\omega}\Delta_t) + (1 - \cos(\tilde{\omega}\Delta_t))^2 - (1 - \cos(\tilde{\omega}\Delta_t)))} \approx \frac{2\pi\tilde{F}\Delta_t}{\sqrt{3}} \approx 3.62579\tilde{F}\Delta_t.$$

The above approximation is possible being in practice $\tilde{\omega}\Delta_t \ll 1$, for the temporal integration to be possible.

The Hannan–Quinn information criterion

As it will be outlined in the sequel, the smoothing constant is the minimiser of a penalized Log-Likelihood function, estimated on a model of the class SARIMA applied to both the original and filtered time series $\hat{\xi}_t$. The objective function chosen belongs to the class of the information criterion, i.e. the Hannan-Quinn criterion.²³ This order selector has been constructed from the law of iterated algorithm and shows a penalty function growing at a very slow rate, as the samples size increases. It is defined as follows:

$$HQc = \log L(\hat{G} | X_t) + 2k \log(\log(T)),$$

being $\hat{G} \equiv (f, F, q, Q, \sigma)$ the vector of the estimated SARIMA parameters of (??) whereas L is the Likelihood function.

By estimating the log-likelihood function $\log L(\hat{G} | X_t)$ with $(\hat{\sigma}^2 | \hat{G})$ and writing the penalty term as $\rho = 2k \log(\log(T))$, the Hannan Quinn information criterion can be written as

$$HQc = f(\hat{\sigma}_o^2, \rho(G)) | X_t^o,$$

where also the subscript (o) , stands for "observed".

The standard identification procedure for the selection of the best SARIMA model order under this information criterion, follows the general rule as the other Information Criteria and is denoted as Minimum Information Criterion Estimation (MICE). In essence, it is based on the minimization of the information criterion itself. Here, the model M_0 minimizing the HQ criterion, is the winning one, that is:

$$M_0^o : (\hat{G}) = \arg \min_{\hat{G}} f(\hat{\sigma}^2, G | X_t^o), \tag{8}$$

with Γ as defined in (2).

The method

The proposed method can be regarded as an optimal averaging procedure under a penalized log-likelihood function. "Properly" averaging a signal as a noise reduction method is consistent to the principle that, if repeated measures are conducted many times on a given signal, part will tend to accumulate but the noise will be irregular and tend to cancel itself. This result is connected to the fact that the mean standard deviation of N measurements is smaller by

a factor of N than the standard deviation of a single measurement. Practically, if one compute the average of “many” samples of a noisy signal the noise-induced random fluctuations tend to die out and the signal is therefore stronger.

In more details, when one records the signal x_t^o (t times (the sampling rate) over a predefined time interval, say τ , being $\tau \leq T$, he can compute l averages each of length $\frac{l}{\tau}$. Two critical points arise at this time, i.e. how many data points should be averaged and in which manner. The latter question has been already answered (Section 3) whereas the former – which is strictly related with the smoothing parameter λ – controls the number of observations of x_t^o to be averaged (and to what extent), so that the noise which randomly fluctuate with equal probability above or below the “true” level x_t^u tends to cancel out while the signal builds up.

The λ parameter is chosen as the minimizer of (8) once a “great” number of candidates $\lambda_j; j=1, \dots, \Lambda$ is tried. Therefore, (8) is modified as

$$M_0^\xi = (\hat{G}, \hat{\lambda}) = \arg \min_{\hat{G}, \hat{\lambda}} f(\hat{\sigma}^2(\varepsilon^o), \mathcal{G} | \xi_t(\lambda)), \tag{9}$$

being $\xi_t = f(x^o(\lambda))$.

What presented in (9) is a slightly modified version of the Hannan Quinn criterion, called here *mHQ* (which will be justified below), whose minimum value delivers the optimal smoothing parameter λ^* .

From the last equation, it is clear that both the optimal model and the optimal λ are simultaneously estimated. However, while the *ML* parameters’ estimation is conducted on the filtered versions ξ_t of the observed signal, the variance parameter is estimated using the observed data x^o . Such an operation is possible assuming $\mathbb{E}[\varepsilon_t] = 0$ and serves the purpose to “force” the filtered output to be “not too far” from the observed signal x_t^o . In practice, the Information Criterion selects an “optimal” model – built on a λ^* -filtered version of the data (which therefore are used to build the *ML* estimations of the *SARIMA* parameters) whose goodness of fit is computed with respect the observed (noisy) signal x_t^o . In other words, *mHQ* criterion has a stronger penalty than the standard one. Such an additional weight (the penalty associated with the residual variance computed on x^o) is proportional to $|\sigma_o^2 - \sigma^2(\xi_t)|$ and serves the purpose to prevent the H-Q criterion from choosing λ values that can lead to over smoothing. In fact, the greater the value of λ the greater the distance between x_t^o and ξ_t and therefore the greater the absolute values of the difference $|\sigma_o^2 - \sigma^2(\xi_t)|$. When no benefit can be obtained by applying the filter, we have $M^\xi(HQ) \geq M_0$ therefore x^o and ξ coincide and $|\sigma_o^2 - \sigma^2(\xi_t)| = 0$, i.e. the best *SARIMA* model is built on x_t^o and its order selection follows the MICE rule according to equation (8). On the contrary, when a benefit from filtering is detected – i.e. $mHQ < HQ$ – therefore the best model is the one built on the filtered time series according to a λ value (λ^*) at which the information criterion (9) is minimized and its order selection follows the MICE rule. Again, such a conclusion is justified as the error term is “average out” by the filter. At this point, we have two different *SARIMA* models fitting x^o and ξ_t and generating two error vectors.

Recalling (3), and consistently with the notation so far used, it will be

$$\varepsilon_t^o = x_t^o - G^o \varepsilon_{t-1}; \quad x^o = 1, \dots, T; \quad \varepsilon_0 = 0. \tag{10}$$

and ε_t^ξ found in a similar way, replacing the superscript identified by the letter o with the Greek letter ξ .

Moreover, by virtue of Granger theorem,²⁴ it is always possible for an estimated *SARIMA* model order to depart from the “true” one when the input is affected by white noise components. In such a case, the CSS estimates resulting from the function $\log Lik(\hat{\Gamma} | x^o)$ can be regarded as converging to the pseudo-true value of the parameters’ vector and therefore the order detected by means of the *mHQ* – which employs the function $\log Lik(\hat{\Gamma} | x_t)$ for the estimation of the parameters – criterion will be closer to the “true” model order.

Signal improvement assessment

At this point, it is clear that for $\Gamma^0 = \Gamma(\xi)$ the lower bound in the gain delivered by the procedure is attained (i.e. no gain and $\lambda = 1$). In this section, a simple way for assessing the performances of the filter (if delivered) is presented. At this point it might be worthwhile emphasizing that the method presented in this paper is not designed to improve the forecasting performance of *SARIMA* models. In fact, the observations coming from the future, under an unchanged data generation mechanism, will be affected, in average, by the same noise components affecting the previous observations. However, projecting the filtered observations into the future is a useful exercise for at least two reasons: i) to verify that the forecasting performances delivered by the two *SARIMA* models are “not too far” from each other (meaning that the ξ_t is a “bona fide” version of the original one; ii) in many cases, improvements in the forecasting performances have been observed. This can be explained by the fact that the order selection and estimation procedures tend to work better under less noisy signals.

With that said, the signal-to-noise ratio (SNR) has been used as an indicator of the quality of the filtered signal. It is defined as the ratio of the power of the signal (the meaningful information) to the power of the noise (the unwanted portion of the signal). In its “pure” form it is

$$SNR = \frac{P_{signal}}{P_{noise}}, \quad P \text{ being the average power. In the framework at hand, both the power levels are estimated by their respective variances } \frac{\hat{\sigma}_{\xi_t}^2}{\hat{\sigma}_{\varepsilon_t}^2},$$

being the error terms obtained using (10). Practically, by comparing

$$SNR^o = \frac{\hat{\sigma}_{x_t^o}^2}{\hat{\sigma}_{\varepsilon_t^o}^2} \text{ with } SNR^\xi = \frac{\hat{\sigma}_{\xi_t}^2}{\hat{\sigma}_{\varepsilon_t^\xi}^2}, \text{ one is able to assess how much of}$$

the unwanted components has been removed under λ^* .

The algorithm

In what follows, the method is presented in the form of a step-by-step algorithm. In particular, steps 2-6 are conducted on the training set whereas the remaining 7-8 use the test set.

1. The time series under investigation is split in two sets: training and test;
- A grid of tentative smoothing constant $\lambda_j; j=1, \dots, \Lambda \in \mathcal{L}$ is built;

A SARIMA model is estimated on the original time series x^o using the H-Q criterion;

Λ SARIMA models are estimated on Λ filtered time series $\xi_j^{(\lambda)} = x^o(\lambda_j); j = 1, \dots, \Lambda$;

If $mHQ \geq HQ$ stop and use the unfiltered time series otherwise proceed;

The λ^* smoothing constant associated with the winner model (according to the modified H-Q criterion mHQ) is then chosen;

Horizon $h; h = 1, \dots, H$ forecast mean square errors are computed for both the filtered and unfiltered series;

Both the statistics are compared for consistency.

Clearly, if one is not interested in forecast comparison – e.g. to speed up the computations and have faster outcomes – then steps 1, 7 and 8 can be skipped.

Empirical experiment

The experiment presented in this paragraph is aimed at showing the capabilities of the proposed method. A time series – collected by the popular search engine Google and recovered in an “ad hoc” repository called Google Trends – will be employed. It refers to the relative number of Italian Internet users which inputted the keyword “caviar” (caviale, in Italian) into the search engine Google, between January 2004 and June 2018. The data have been download on June 6th 2018 (h 03:29 PM GMT). Time series length is 174, split in training set – ending in December 2014 (132 data) – and test set (42 data), ending in June 2018. The last datum has been purposely left in the data set, for it is an even more unstable one (in fact, it is supported only by a small portion of the month of reference). A visual inspection of the series, depicted in Figure 1, shows many irregularities, a pronounced, non stationary seasonality as well as a non stationary trend. Eteroschedasticity is also an issue. Such a behavior might be considered typical for a luxury good especially when observed, as in the present case, for a long period of time.

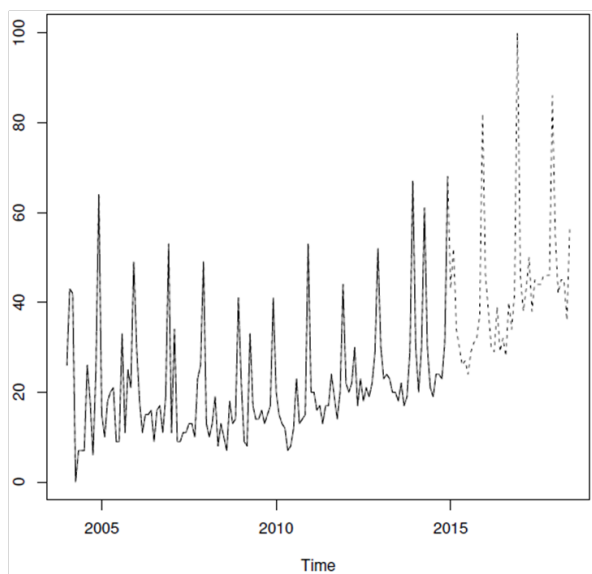


Figure 1 Relative number of the keyword “caviar” searched by the Italian Internet users. Training set (continuous line) and test set (dashed line).

The whiteness of the residuals has been tested using a test of the type “Portmanteau”, i.e. the Ljung-Box test,²⁵ which is designed to account for the sum of the autocorrelations (denoted by the Greek letter ρ) until a predetermined, arbitrary lag K , i.e.

Publication

$$Q = T(T+2) \sum_{k=1}^K \frac{\hat{\rho}_k^2}{T-K} \underset{H_0}{\sim} \chi_{(K)}^2; \quad Q^{rej} > \chi_{(1-\alpha, K)}^2.$$

Here Q^{rej} is, for significance level α , the critical region for rejection of the hypothesis of randomness under the χ^2 distribution.

The estimated SARIMA order in the case of x^o is $(1,1,0) \times (2,0,0)$ with an HQ criterion of 4.97. In the case of ξ_t , the estimated SARIMA order is $(0,0,1) \times (0,1,1)$, for a modified HQ criterion equal to 4.05. Therefore, the filter is applied. The effects of its application are presented in Figure 2, where a more regular signal can be noticed. Such an impression is confirmed by looking at the SNRs, which is equal to 6.88 and 7.06 respectively for the unfiltered and filtered time series. The optimal smoothing constant $\lambda^* = .805$, has been obtained at iteration number 82 (see Figure 3), using a grid of 121 λ values (equally spaced from 0.400 to 1; incremental step=0.005). The residuals (computed in both the cases on the original time series) pass the ‘Portmanteau’ test (computed for the lags 1–24) only in the case of ξ_t (p value = 0.382), whereas for the raw signal we have p value ≈ 0 . Such a conclusion is consistent with the patterns shown by the empirical autocorrelation functions (see Figure 4) for x^o (graph a) and ξ_t (graph b). In the same graph, the residual density distributions for x^o (graph c) and ξ_t (graph d) confirm a better behavior when the proposed method is applied. Finally, in Figure 5, the prediction performances obtained using ξ_t appears to be slightly better until lag 8.

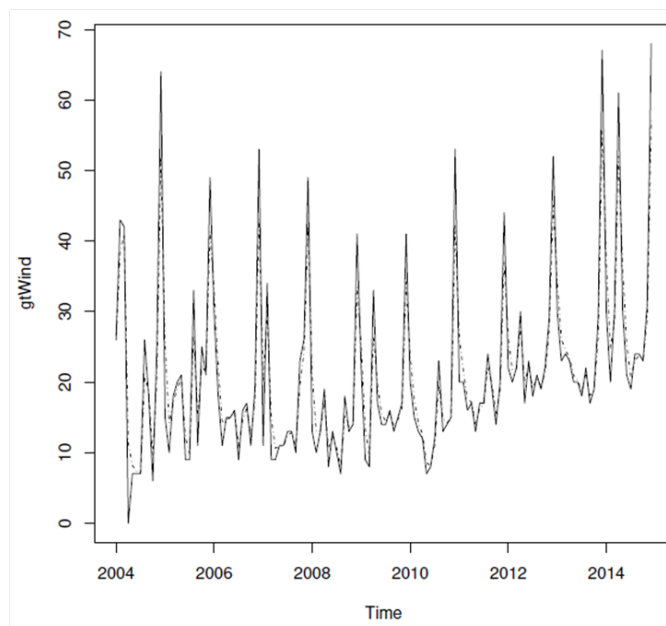


Figure 2 Unfiltered (continuous line) and filtered (dashed line) time series.

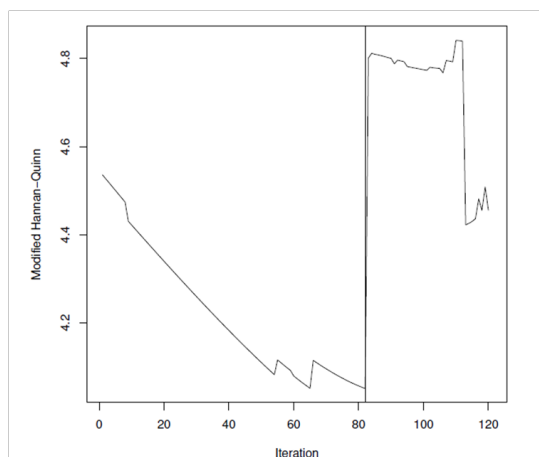


Figure 3 Values of the modified Hanna–Quinn information criterion for each iteration (attempted λ value). The vertical line shows the iteration number at which the minimum mHQ value is found.

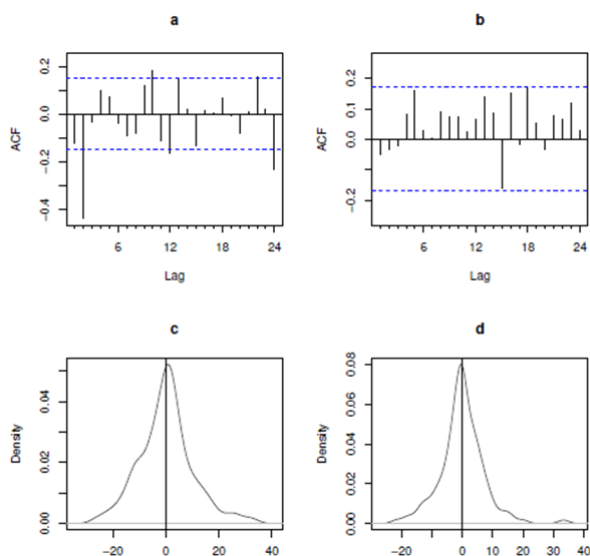


Figure 4 Autocorrelation functions for x_t^o (graph a), ξ_t (graph b) and residual density for x_t^o (graph c) and ξ_t (graph d).

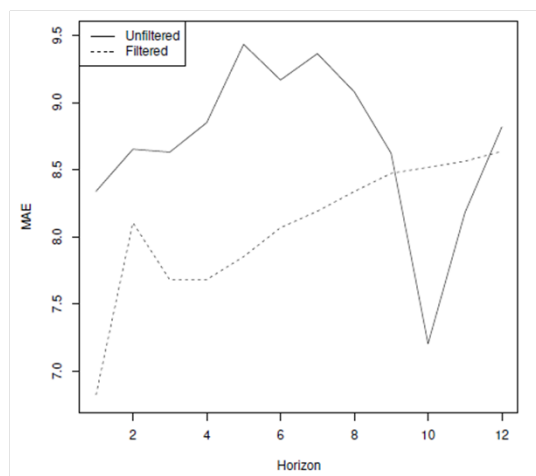


Figure 5 Mean Absolute Error of the predictions obtained from the two SARIMA models applied on the raw time series (continuous line) and the filtered time series (dashed line). Time horizon 1–12.

Conclusion

One effective way to reduce noise components affecting our time series is to use a First Order Discrete Time Infinite Impulse Response Filter. As it is well known, its filtering performances are critically dependent on the smoothing constant, which therefore has to be carefully fine-tuned. The method proposed in this paper provides the practitioner with a useful – and reasonably fast – tool to filter out such unwanted components. However, that the objective function used – based on the Hannan–Quinn information criterion – has given good results, does not imply that is the best possible one. For example, with small data set it could be worth considering different information criteria, specifically designed for such a case, e.g. the second-order AIC.²⁶ Finally, different stochastic models can be always considered, according to the underlying data generating process. For instance, whenever the time series shows memory characteristics of the type long range, the proposed procedure can be associated to a suitable model (e.g. of the class Autoregressive Fractional Integrated Moving Average).

Acknowledgements

None.

Conflict of interest

Author declares that there is no conflict of interest.

References

1. Richard T O'Connell, Anne B Koehler. Forecasting, time series, and regression: An applied approach, volume 4. USA: South-Western Pub; 2005.
2. Ruey S Tsay. *Analysis of Financial time series*. Volume 543. USA: John Wiley & Sons; 2005.
3. Harold Wayne Sorenson. Kalman ltering: theory and application. *IEEE*. 1985.
4. Phillip G Gould, Anne B Koehler, J Keith Ord, et al. Forecasting time series with multiple seasonal patterns. *European Journal of Operational Research*. 2008;191(1):207–222.
5. David L Donoho. De-noising by soft-thresholding. *IEEE transactions on information theory*. 1995;41(3):613–627.
6. Albert Cohen, Ingrid Daubechies, Bjorn Jawerth, et al. Multiresolution analysis, wavelets and fast algorithms on an interval. *Comptes Rendus de l'Académie des Sciences*. 1993;316(5):417–421.
7. Holger Kantz, Thomas Schreiber. Nonlinear time series analysis, volume 7. USA: Cambridge university press; 2004.
8. S Makridakis. Forecasting: Methods and applications. 1998.
9. Torsten kohler, Dirk Lorenz. *A comparison of denoising methods for one dimensional time series*. Bremen: Germany, University of Bremen; 2005.
10. George EP Box, Gwilym M Jenkins, Gregory C Reinsel, et al. *Time series analysis: forecasting and control*. USA: John Wiley & Sons; 2015.
11. Peter J Brockwell, Richard A Davis. Time series: theory and methods. *Springer Science & Business Media*. 2013.
12. JH Stock, MW Watson. *Introduction to econometrics*. Boston: Addison-wesley; 2003.
13. MJ Evanich, AO Newberry, LD Partridge. Some limitations on the removal of periodic noise by averaging. *Journal of Applied Physiology*. 33(4):536–541;1972.

14. CC Holt. Forecasting trends and seasonals by exponentially weighted averages. carnegie institute of technology. *Technical report, Pittsburgh ONR memorandum*. 1957.
15. Adrien Cahuzac, Jerome Boudet, Pierre Borgnat, et al. Smoothing algorithms for mean-ow extraction in large-eddy simulation of complex turbulent ows. *Physics of fluids*. 2010;22(12):125104.
16. Layth C Alwan, Harry V Roberts. Time series modelling for statistical process control. *Statistical Process Control in Automated Manufacturing*. 1988;15:45–65.
17. J Stuart Hunter. The exponentially weighted moving average. *Journal of quality technology*. 1986;18(4):203–210.
18. Gery Casiez, Nicolas Roussel, Daniel Vogel. 1 euro Filter: a simple speed-based low-pass Filter for noisy input in interactive systems. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. *ACM*. 2012;2527–2530.
19. Adrien Cahuzac, Jerome Boudet, Pierre Borgnat, et al. Smoothing algorithms for mean-ow extraction in large-eddy simulation of complex turbulent ows. *Physics of fluids*. 2010;22(12):125104.
20. Rob Hyndman, Anne B Koehler, J Keith Ord, et al. Forecasting with exponential smoothing: the state space approach. *Springer Science & Business Media*. 2008.
21. Fat Codrut, A Maria, Dezsi Eva, et al. Exchange-rates forecasting: Exponential smoothing techniques and arima models. *Annals of Faculty of Economics*. 2011;1(1):499–508.
22. Athanasios Papoulis. *Signal analysis*. volume 191. New York: McGraw-Hill; 1977.
23. Edward J Hannan, Barry G Quinn. The determination of the order of an autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1979; 190–195.
24. Clive WJ Granger. *Essays in econometrics: collected papers of Clive WJ Granger*. Volume 32. USA: Cambridge University Press; 2001.
25. Greta M Ljung, George EP Box. On a measure of lack of t in time series models. *Biometrika*. 1978;65(2):297–303.
26. Kenneth P Burnham, David R Anderson. Model selection and multimodel inference: a practical information-theoretic approach. *Springer Science & Business Media*. 2003.