Singular Spectrum Analysis for Value at Risk in Stochastic Volatility models

Josu Arteche and Javier García-Enríquez*†

University of the Basque Country UPV/EHU
Avda. Lehendakari Aguirre 83, 48015 Bilbao, Spain

Revised: 12th February 2021

Abstract

Estimation of the Value at risk (VaR) requires prediction of the future volatility. Whereas this is a simple task in ARCH and related models, it becomes much more complicated in Stochastic Volatility (SV) processes where the volatility is a function of a latent variable that is not observable. In-sample (present and past values) and out-of-sample (future values) prediction of that unobservable variable are thus necessary. This paper proposes Singular Spectrum Analysis (SSA), which is a fully non-parametric technique that can be used for both purposes. A combination of traditional forecasting techniques and SSA is also considered to estimate the VaR. Their performance is assessed in an extensive Monte Carlo and with an application to a daily series of SP500 returns.

*Corresponding author: Javier García-Enríquez. E-mail address: javier.garcia@ehu.es

†Research supported by the Spanish Ministry of Science and Innovation and ERDF grant ECO2016-76884-P, National Research Agency grant PID2019-105183GB-I00 and UPV/EHU Econometrics Research Group (Basque Government grant IT1359-19). The comments provided by one referee are gratefully acknowledged.

1 Introduction

Prediction is very difficult, especially if it's about the future.

(usually attributed to the Nobel laureate in Physics Neils Bohr)

Cambridge Dictionary defines prediction as *a statement about what you think will happen in the future*. In a time series setting that prediction implies estimation of the observation at time $t+h$ with the information provided upto time t, where h is the horizon of prediction. The definition in Cambridge Dictionary asserts that prediction always refers to the future entailing $h > 0$ and thus the previous quote might be considered just a rhetorical figure. However, the quote makes sense when dealing with latent variables where prediction may be *in-sample* ($h \leq 0$) or *out-of-sample* ($h > 0$). The former implies estimation of past and present values of a latent variable using all the information contained within the sample whereas the latter refers to estimation of future values at time $t + h$ for $h > 0$ using observations upto time t. Thus, *in-sample* prediction with $h \leq 0$ does not refer to the future but it implies signal extraction where the latent variable is the signal that needs to be extracted from the available information composed of the observables. More difficult, as stated in the quote, is the prediction of future values, especially when dealing with latent variables because no past or current observations of the variable to be predicted are available.

Stochastic Volatility (SV) models are typical examples of this kind of situations, where the volatility is the latent variable that needs in-sample and out-of-sample prediction to assess the risk premium (in-sample) and the Value at Risk (VaR) (out-of-sample). The former requires extraction of a latent signal, and there exist nowadays many techniques for that purpose. One of them is the fully nonparametric Singular Spectrum Analysis (SSA) proposed recently by Arteche and García-Enríquez (2017), who show the advantages of SSA over competitors. But the main focus of this paper is out-of-sample prediction and estimation of the VaR, where SV models are of especial value (see González-Rivera et al., 2004). This is currently of great importance because financial institutions are required to hold regulatory capital based on their VaR forecasts. We propose a combination of SSA for in-sample volatility prediction with some out-of-sample forecasting techniques to estimate the VaR in SV models. The methods for out-of-sample prediction considered range from traditional strategies such as ARIMA models to more sophisticated techniques such as the extension of SSA for out-of-sample prediction, never before used for the estimation of the VaR.

The rest of the paper is structured as follows. SV models and the formal definition of VaR are described in Section 2. Section 3 details the SSA technique proposed for signal extraction or in-sample prediction of the volatility. These in-sample predictions are used as the base for out-of-sample prediction of the volatility as described in Section 4, which pays special attention to its use for VaR estimation. The performance of SSA and other prediction techniques for VaR estimation in SV models is analysed in a Monte Carlo in Section 5. Finally, Section 6 applies these techniques to estimate the VaR in a daily series of SP500 returns, presenting some backtesting results.

2 Value at Risk in Stochastic Volatility models

We consider SV models of the form

$$z_t = \sigma_t \varepsilon_t \ , \tag{1}$$

where $\sigma_t = \sigma \exp(v_t/2)$ is the conditional volatility for σ a positive constant scale factor, v_t is the volatility component and $\varepsilon_t \sim iid(0,1)$ (see Taylor, 1986). The series z_t can be residuals after extracting the dynamics of the conditional mean. In this context estimation of the conditional volatility σ_{t+h} is interesting to assess the risk premium ($h \leq 0$) or to estimate the VaR ($h > 0$). Linearising by taking logs of the squares gives a signal plus noise process of the form

$$y_t = \log z_t^2 = \mu + v_t + \xi_t, \tag{2}$$

with $\mu = \log \sigma^2 + E \log \varepsilon_t^2$, and where the added noise $\xi_t = \log \varepsilon_t^2 - E \log \varepsilon_t^2$ is $i.i.d.$ with zero mean and variance σ_ξ^2. For example, if $\varepsilon_t \sim N(0,1)$ then ξ_t is a centred $\log \chi_1^2$ variable with $E \log \varepsilon_t^2 = -1.27$ and $\sigma_\xi^2 = \pi^2/2$. However the Gaussianity in financial time series is often questioned and some leptokurtic distribution can be used instead to adapt to the excess of kurtosis usually found in those series. The Student´s t distribution is one popular alternative for ε_t. If $\varepsilon_t \sim t_v$ then $E \log \varepsilon_t^2 = \psi(1/2) - \psi(v/2) + \log v$ and $\sigma_\xi^2 = \psi'(1/2) + \psi'(v/2) =$

$\pi^2/2 + \psi'(v/2)$, where ψ and ψ' are the digamma and trigamma functions respectively. It is thus interesting to propose distribution-free forecasting techniques, in the sense that they do not rely on the particular distribution of v_t or ξ_t. Uncorrelation between v_t and ξ_s at all leads and lags is assumed hereafter. Note that this does not preclude the possible existence of leverage in the form of correlation between v_t and ε_s. In fact, as long as the joint distribution of v_t and ε_s is symmetric around the origin (e.g., Gaussian, Student's t or a Generalized Error Distribution among others) the possible correlation between the two does not preclude the absence of correlation between v_t and ξ_s (see Harvey et al., 1994).

The Value at Risk (VaR) is defined as the quantile in the lower tail of the distribution of a future return, conditional on the available information set I_n. More explicitly, the h-horizon VaR for $h > 0$ and a confidence $1 - \alpha$ is defined as the α-quantile of the conditional future return distribution, i.e.

$$VaR_{n+h|n}(\alpha) = \inf_x \{x \in \mathbb{R} : Pr(z_{n+h} \leq x | I_n) \geq \alpha\},$$

which for SV models is estimated as $\widehat{VaR}_{n+h|n}(\alpha) = \hat{\sigma}_{n+h|n} q_\varepsilon(\alpha)$, where $Prob(\varepsilon_t < q_\varepsilon(\alpha)) = \alpha$ and $\hat{\sigma}_{n+h|n}$ is the out-of-sample prediction of the volatility with horizon of prediction h based on the information set upto period n. $\hat{\sigma}_{n+h|n}$ can be obtained via out-of-sample predictions of v_{n+h}, which are then plugged into the formula for σ_t to obtain the estimates of the out-of-sample volatilities $\hat{\sigma}_{n+h|n} = \hat{\sigma} \exp(\hat{v}_{n+h|n}/2)$. In-sample predictions of the v_t and σ_t, that is $\hat{v}_{t|n}$ and $\hat{\sigma}_{t|n} = \hat{\sigma} \exp(\hat{v}_{t|n}/2)$ for $t = 1, ..., n$, are also useful for several purposes. First, the constant σ can be estimated as the square root of the sample variance of $z_t \exp(-\hat{v}_{t|n}/2)$, $t = 1, ..., n$ (see equation (4) below). Second, $q_\varepsilon(\alpha)$ can be obtained as empirical quantiles of the Studentised observations $z_t/\hat{\sigma}_{t|n}$, $t = 1, ..., n$, where the series of returns are prefiltered by using in-sample predictions of the volatility in order to approach the behaviour of ϵ and exploit its iid characteristic. This strategy resembles the Filtered Historical Simulation (FHS) in that Studentised observations are used to obtain $q_\varepsilon(\alpha)$ (see Barone-Adesi et al. 1998 or McNeil and Fey, 2000). Third, $\hat{v}_{t|n}$, $t = 1, 2, ..., n$ can be used as the base to predict v_{n+h} with standard prediction techniques, avoiding in that way the distorting effect of the added noise (see for example Soofi and Cao, 2002). More details are given in Section 4. Following Arteche and García-Enríquez (2017), Singular Spectrum Analysis (SSA) is proposed for in-sample estimation of the volatility as described in next section.

3 In-sample prediction: SSA for signal extraction

Singular Spectrum Analysis (SSA) is a relatively novel tool for signal extraction and prediction of a latent variable in a signal plus noise process as that in (2). Although previously used in signal processing and oceanography, its application in the analysis of time series was generalised by Broomhead and King (1986) who applied it in nonlinear dynamics. Finally, the book by Golyandina et al. (2001) extended its interest in wider areas of time series analysis. The procedure makes use of the different spectral behaviour of signal and noise to identify and estimate the signal. This makes it especially useful for in-sample prediction of the volatility component v_t in SV models, where the noise satisfies $\xi_t \sim iid(0, \sigma_\xi^2)$ with flat spectrum. Using the complete uncorrelation between ξ_t and v_t, all the spectral structure in y_t is thus due to v_t, which permits identification of the volatility component. SSA is then based on decomposing the original series into a sum of components so that each of those components is characterised by a particular spectral behaviour, displayed in the periodogram. The signal can then be estimated by reconstructing the original series using only those components that share a spectral behaviour similar to that of the latent signal. For example, a signal with a persistent trend can be estimated by selecting those components with spectral concentration around the origin, with no need to know whether that trend is deterministic or stochastic, stationary or non-stationary.

The detailed procedure consists of the following steps:

- **Step 1**: Let $y_t^* = y_t - \hat{\mu}$ be the centred series of observations for $t = 1, 2, ..., n$ (we use $\hat{\mu} = \sum y_t/n$ but other options are also possible, see Arteche, 2015). Construct the trajectory matrix $Y = [Y_1 : ... : Y_K]$ for $Y_j = (y_j^*, ..., y_{j+L-1}^*)'$ where $1 < L < n$ is called the window length and $K = n - L + 1$. By definition Y is a Hankel matrix and we assume with no loss of generality that $L \le K$.

- **Step 2**: Apply the Singular Value Decomposition (SVD) to Y

$$Y = \sum_{j \in J} \sqrt{\mu_j} U_j V_j' \ , \ \ V_j = \frac{1}{\sqrt{\mu_j}} Y' U_j \ , \ \ J = \{j \ \text{ such that } \mu_j > 0\},$$

where μ_j and U_j are the j-th eigenvalue and eigenvector respectively of YY'. Each U_j is an $L \times 1$ vector known as Empirical Orthogonal Function (EOF).

- **Step 3**: Reconstruction. Select a subgroup $\gimel$ of SVD components and form the matrix

$$Y_\gimel = \sum_{j \in \gimel} \sqrt{\mu_j} U_j V_j'.$$

  $\gimel$ contains the SVD components with EOFs sharing the same spectral characteristics as the latent signal.

- **Step 4**: Estimate v_t, $t = 1, ...n$, by Hankelization of the matrix $Y_\gimel$ as

$$\hat{v}_{t|n}^{ssa} = \begin{cases} \dfrac{1}{t} \displaystyle\sum_{l=1}^{t} c_{l,t-l+1} & 1 \leq t \leq L, \\ \dfrac{1}{L} \displaystyle\sum_{l=1}^{L} c_{l,t-l+1} & L < t \leq K, \\ \dfrac{1}{n-t} \displaystyle\sum_{l=t-K+1}^{L} c_{l,t-l+1} & K < t \leq n, \end{cases}$$

  where $c_{j,k}$ is the (j,k)-th element of the matrix $Y_\gimel$ (see Vautard et al. 1992).

The procedure requires the intervention of the user in Steps 1 and 3 to select L and $\gimel$. Golyandina (2010) suggests a window length L multiple of the periodicity of the series and close to but smaller than half the sample size to obtain minimal errors in many situations. See also Golyandina et al. (2001). We follow this suggestion in the Monte Carlo and the empirical application below. For the selection of $\gimel$, Arteche and García-Enríquez (2017) suggest to choose a set containing those elements with EOFs with spectral concentration, measured through the periodogram of the observables, similar to the signal. This is justified because under uncorrelation of v_t and ξ_s $\forall t, s$, the spectral density function of y_t ($f_y(\lambda)$) is the sum of the spectral densities of v_t ($f_v(\lambda)$) and ξ_t ($f_\xi(\lambda)$). But $f_\xi(\lambda)$ is constant and then $f_y(\lambda)$ is just a shifted version of $f_v(\lambda)$. Thus, y_t inherits the spectral characteristics of v_t and can be used to select the components in $\gimel$. In particular, for low frequency behaviours as those considered in the Monte Carlo below, $\gimel$ is selected to contain those components with EOFs showing relative periodogram concentration as large as that in the observable series y_t at the k closer to zero Fourier frequencies where k is a user-chosen value. See Arteche and García-Enríquez (2017) for details and recommendations on how to select k.

The decomposition in Step 2 is based on YY', forming what is usually known as Basic SSA. Other alternatives are the Toeplitz SSA (Vautard and Ghil 1989) who use the sample lagged covariance matrix instead of YY' in a stationary context and the Circulant SSA (Bógalo et al. 2021) based on a circulant alternative matrix of second moments. This recent alternative seems promising because it identifies eigenvalues with the spectral density, which may be used for selection of components in the reconstruction Step 3, but further analysis is necessary. Moreover all these techniques offer similar empirical performance (Bógalo et al. 2021) and thus we stick to the more popular Basic SSA.

4 Out-of-sample prediction: Forecasting techniques

Taking into account the latent character of v_t, two general approaches can be considered to predict v_{n+h}. First, noting that the optimal predictor of v_{n+h} in (2) coincides with the optimal predictor of $y_{n+h} - \mu$, traditional forecasting techniques can be implemented in the centred sample of observables $y_1, ..., y_n$ to predict $y_{n+h} - \mu$ and use this prediction as the forecast of v_{n+h}. However, the added noise may cause significant differences between the empirical prediction and v_{n+h} (see for example Soofi and Cao, 2002). To avoid that potentially distorting effect of the added noise, the second approach consists in predicting v_{n+h} by applying the forecasting strategies on the estimated v_t, $t = 1, ..., n$ obtained by the application of some signal extraction technique. Following Arteche and García-Enríquez (2017) we use SSA as described in the previous section. Once we get rid of the noise the estimated signal is predicted. In this case we avoid the effect of the added noise but instead the prior estimation of v_t may significantly affect the performance of the prediction of v_{n+h}.

We consider five different techniques to predict v_{n+h}, $h = 1, 2, ...$ The first one is an extension of the SSA for out-of-sample forecasting of v_{n+h}. This is a quite novel technique never before used for estimation of the VaR. The other four are more traditional techniques, which are applied on the series of centred observables y_t^* and on the signal extracted using SSA, $\hat{v}_{t|n}^{ssa}$. First, two different techniques, embedded in the popular Box-Jenkins methodology, are considered. They make use of fitted AR and $ARIMA$ processes. Then Exponential Smoothing and Neural Networks as implemented in the package *forecast* in R are also implemented. The application of these four techniques on the series of observables y_t^* and on the signal extracted using SSA $\hat{v}_{t|n}^{ssa}$ permits us to assess if the added noise exerts a negative effect in the prediction of the signal as advocated by Soofi and Cao (2002) or if, on the contrary, the error generated in the extraction of the signal is of such a magnitude that using y_t^* is a better option. In order to distinguish notationally both strategies, a hat is used to denote predictions that have been obtained by applying the corresponding technique to $\hat{v}_{t|n}^{ssa}$ whereas a tilde is used if it has been applied to the series of observables y_t^*. A total of nine different forecasting strategies are thus considered.

4.1 SSA forecasting

The SSA algorithm can be extended to predict future values of the latent signal using a linear recurrence expression for v_t that does not depend on the actual structure of the latent variable. The SSA prediction of v_{n+h} is defined as

$$\hat{v}_{n+h|n}^{ssa} = R'\hat{V}$$

for $\hat{V} = (\hat{v}_{n+h-L+1|n}^{ssa}, ..., \hat{v}_{n+h-1|n}^{ssa})'$, $R = (1 - v^2)^{-1} \sum_{j=\beth} \pi_j U_j^{\Delta}$ for $v^2 = \sum_{j=\beth} \pi_j^2$, π_j is the L-th element of U_j and U_j^{Δ} denotes the first $L-1$ elements of U_j. For $h > 1$ the predictions are obtained recursively (see Golyandina et al. 2001).

The application of SSA for forecasting is recently gaining attention and has been discussed and implemented in real time series by Lisi and Medio (1997), Hassani et al. (2009), Beneki et al. (2012), de Carvalho et al. (2012), Hassani et al. (2013), Silva and Hassani (2015), Khan and Poskitt (2017), Papailias and Thomakos (2017), among others. Unlike all this extant literature centred on prediction of observables, we focus on prediction of a latent variable and its consequent use for VaR estimation in SV models. An additional contribution is the analysis of its performance in an extensive Monte Carlo in Section 5, not relying only on its performance in a single time series as all the mentioned papers do.

4.2 Box-Jenkins methodology

One of the main benefits of the Box-Jenkins methodology is that it provides a simple and effective tool to predict a time series exploiting the linear dependence existing within the sample. Mathematically tractable ARIMA models are proposed to capture such dependence and to extrapolate the behaviour of the sample into the future. Considering that the main predictive capacity of ARIMA models lies in the AR part, we consider two different strategies for prediction:

- Fit an $AR(p)$ model, either to $\hat{v}^{ssa}_{t|n}$ or to y^*_t, $t = 1, ..., n$, with p selected by minimizing the AIC and use the estimated model to obtain the predictions $\hat{v}^{ar}_{n+h|n}$ and $\tilde{v}^{ar}_{n+h|n}$.
- Fit an $ARIMA(p, d, q)$ with d selected by KPSS and p, q by minimizing the AIC and use the estimated model to get $\hat{v}^{arima}_{n+h|n}$ if $\hat{v}^{ssa}_{t|n}$ is used and $\tilde{v}^{arima}_{n+h|n}$ if the model is applied on y^*_t, $t = 1, ..., n$.

The main difference between both strategies is that the second one includes the possibility of unit roots, which may be relevant for predicting strong dependent series, whereas the first one proposes to capture such dependence with a large enough p.

4.3 Exponential Smoothing (ETS)

While linear exponential smoothing models are all special cases of ARIMA, non-linear exponential smoothing models are far more general and afford different types of behaviour. Additive and multiplicative, possibly damped components are here considered and the specification that is used for forecasting is selected by minimizing the AIC. See Hyndman et al. (2002) or Hyndman and Khandakar (2008) for more details. Using $\hat{v}^{ssa}_{t|n}$ or y^*_t as input variables leads to $\hat{v}^{ets}_{n+h|n}$ and $\tilde{v}^{ets}_{n+h|n}$ respectively.

4.4 Autoregressive Neural Networks (NN)

Neural networks are universal approximations of non-linear functions. In this case, a feedforward neural network is considered with p lagged values of the variable to be predicted as inputs and a single hidden layer with k hidden nodes. The mathematical expression of this neural network is then

$$x_t = \beta_0 + \sum_{j=1}^{k} \beta_j \psi(x_{t-1}, .., x_{t-p}; \gamma_j).$$

The logistic function is used as activation function such that

$$\psi(x_{t-1},..,x_{t-p};\gamma_j) = \left[1+\exp\left(-\gamma_{j0}+\sum_{i=1}^{p}\gamma_{ji}x_{t-i}\right)\right]^{-1}.$$

The number of lags p is selected by AIC when applied to a linear $AR(p)$. The number of hidden nodes is selected as $k = (p+1)/2$ rounded to the nearest integer. The predictions of x_{n+h} are recursively obtained based on this specification. See Hyndman and Athanasopoulos (2014) for a more detailed description of the methodology. Replacing x_t by $\hat{v}^{ssa}_{t|n}$ or y^*_t we get $\hat{v}^{nn}_{n+h|n}$ and $\tilde{v}^{nn}_{n+h|n}$ respectively.

5 Finite sample behaviour: VaR evaluation

All the prediction techniques described in the previous section, except the SSA, are implemented using the package *forecast* in *R*. Forecasting with the Kalman Filter based on the $AR(p)$ with p selected by AIC was also considered, but the results were worse than those shown here, and thus are not included (available upon request).

In this Monte Carlo SV models as defined in equations (1) and (2) are generated with $\sigma = 1$ and three different distributions for ε_t: $\mathcal{N}(0,1)$ and t_ν for degrees of freedom $\nu = 4, 8$. The value $\nu = 4$ implies an infinite kurtosis whereas $\nu = 8$ is closer to the values estimated in real financial time series (see for example Liesenfeld and Jung, 2000). Note that all the prediction techniques considered in this Monte Carlo are distribution free in the sense that they do not rely on any particular distribution for their implementation. The signal v_t is chosen to be of the form:

Model 1: $(1-\phi B)v_t = \sigma_1 w_t$ with $\phi = 0.8$.

Model 2: $v_t = \sigma_2(v_{1t} + 2\mu_t)$ where $(1-0.8B)v_{1t} = w_t$ and μ_t is a level shift of the form:

- **a)** Deterministic level shift: $a_t = I_{t>n/4}$ and $\mu_t = a_t - \bar{a}$.
- **b)** Stochastic level shifts: $\mu_t = \sum_{j=1}^{t}\delta_j\eta_j$.

Model 3: $(1-B)^d v_t = \sigma_3 w_t$ for $d = 0.4$.

Model 4: $(1-B)^d v_t = \sigma_4 w_t$ for $d = 0.8$.

where B is the backshift operator ($B^k v_t = v_{t-k}$), w_t and η_j are independent standard normal random variables and δ_j is a sequence of independent Bernoulli random variables taking value 1 with probability p, i.e. $\delta_j \sim B(1,p)$. The variables w_t, η_r and δ_s are mutually independent for all t, r, s.

These models have been previously considered in Arteche and García-Enríquez (2017) to asses the performance of the SSA for signal extraction, showing that SSA performs in general better than other alternatives. The analysis is restricted to these four models with low frequency spectral concentration, either stationary (Models 1 and 3) or non-stationary (Models 2 and 4), because this is the most common situation in economic and financial time series where SV models are employed. For example, the $AR(1)$ in Model 1 was the process that attracted most interest in the origins of SV modeling (see Harvey et al., 1994, among many others). Later, the high persistence found in the volatility of most financial series led many authors to propose a v_t with long memory as in Models 3 and 4 (e.g. Bollerslev and Mikkelsen, 1996, Harvey, 1998, Breidt et al. 1998, Deo and Hurvich, 2001, Arteche, 2004, Hurvich, Moulines and Soulier, 2005, Arteche, 2015). Other authors, however, suggest that the apparent strong persistence in the volatility of financial time series is not due to long memory in v_t but to level shifts as those in Model 2 (Lobato and Savin, 1998, Granger and Hyung, 2004, Perron and Qu, 2010, Qu and Perron, 2013). For the stochastic shift we follow Perron and Qu (2010) and consider $p = 10/n$, which models relatively infrequent shifts that disappear as the sample size increases. Thus, the models considered here cover a wide range of realistic and diverse situations.

Two different Noise to Signal Ratios, defined as $NSR = Var(\xi_t)/Var(v_t)$, are considered: $NSR = \pi^2, 5\pi^2$. In the generated models $Var(\xi_t) = \pi^2/2$ if $\varepsilon_t \sim \mathcal{N}(0,1)$ and $Var(\xi_t) = \pi^2/2 + \psi'(v/2)$ for Student's t_v, where ψ' is the trigamma function. The constants σ_i, $i = 1,2,3,4$ are chosen to adapt to the selected NSR. In Models 1, 2a) and 3 $Var(v_t)$ are population variances whereas in the non-stationary Models 2b) and 4 the population variance is undefined and then $Var(v_t)$ is replaced by the sample variance and the constants σ_3 and σ_4 adjusted in every replication in order to maintain the NSR fixed.

The sample size is $n = 2048$ which is comparable to the size of many financial series as the one analysed in the empirical section. The number of replications is 1000. The SSA is applied using a window length $L = 1008$ (similar values give similar results) in keeping with the comments in the previous section. The selection of $\gimel$ is based on $k = 5$ (values of $k = 10$ and $k = 15$ were also analysed, giving similar results) as explained in Section 3 (see also Arteche and Garcia-Enríquez, 2017).

Once the predictions of v_{n+h} are obtained, the variances of the returns conditional on the volatility component are predicted as

$$\bar{\sigma}^2_{n+h|n} = \hat{\sigma}^2 \exp(\bar{v}_{n+h|n}), \tag{3}$$

where the $\bar{v}_{n+h|n}$ are obtained by one of the forecasting techniques discussed before and the constant σ^2 is estimated with the signal extracted by SSA as

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^{n} z_t^2 \exp(-\hat{v}^{ssa}_{t|n}). \tag{4}$$

This prediction of the conditional variance is used to estimate the h-horizon VaR for a confidence $1 - \alpha$ conditional on the information at time n: $\widehat{VaR}_{n+h|n}(\alpha) = \bar{\sigma}_{n+h|n} q_\varepsilon(\alpha)$, where $Prob(\varepsilon_t < q_\varepsilon(\alpha)) = \alpha$. The quantile $q_\varepsilon(\alpha)$ is calculated as the sample quantile of the Studentised series $z_t/\hat{\sigma}^{ssa}_{t|n}$, $t = 1, ..., n$, where $\hat{\sigma}^{ssa}_{t|n}$ are in-sample predictions of the latent volatility obtained with SSA. Note that SSA is then used for VaR estimation with all the different out-of-sample forecasting techniques here considered, being necessary in two of the steps: in the estimation of the constant σ^2 and to obtain the quantile $q_\varepsilon(\alpha)$.

In order to evaluate the adequacy of the different techniques for VaR estimation, we use two criteria. First, the proportion of VaR violations, that is the proportion of times that $z_{n+h} < \widehat{VaR}_{n+h|n}(\alpha)$, is calculated. The closer this value is to the nominal risk α the better the performance. Tables 1-2 show average distances (in absolute values) of the proportion of violations to the nominal risks 0.05 and 0.01 (multiplied by 100). The second criteria is the magnitude of the exceedance. Not only the number of violations is important, but the magnitude of the exceedance when a violation occurs is also of great relevance for the estimation of the VaR. Undoubtedly a violation with a large loss is more harmful than one with a small exceedance. Tables 3-4 show average losses based on a quadratic loss function defined as

$$L(VaR_{n+h|n}(\alpha)) = \begin{cases} (z_{n+h} - VaR_{n+h|n}(\alpha))^2 & \text{if } z_{n+h} \leq VaR_{n+h|n}(\alpha) \\ 0 & \text{if } z_{n+h} > VaR_{n+h|n}(\alpha) \end{cases} \tag{5}$$

In order to evaluate the factors that may affect the performance of the different forecasting techniques, different averages are presented in Tables 1-4. For example the row starting with $N(0,1)$ shows average distances and losses over 1000 replications, all the models, two NSRs and three horizons of predictions with standard normal innovations. The numbers in the rest of rows are similarly obtained. The following conclusions can be extracted:

- The first question to answer is if the different prediction techniques should be applied on the original series or on the in-sample SSA estimation of the signal. Based on the obtained results the answer is not categorical but it depends on the applied technique. Whereas AR and ETS give smaller distances when applied to the original series, NN and ARIMA tend to offer better results when applied to the extracted signal.

- If we consider only predictions implemented on the estimated signals, SSA is usually the best option, the only exceptions being the average loss for $\alpha = 0.05$, where other techniques have lower losses. If we include also predictions implemented on the centred observables, SSA tends to be the second best, in general only beaten by ETS when applied on the original series, which tends to be the best option.

- Even though the frequencies of violation with leptokurtic innovations are closer to the nominal risk than the frequencies with standard normal innovations, larger kurtoses lead to greater losses.

- Larger NSRs lead to frequencies of violation further from the nominal risk but the losses are smaller. However, this does not imply that larger NSRs produce lower losses. The explanation is more related with the fact that the variance of the signal in the low NSR case is five times larger than the variance with high NSR, affecting directly the quadratic loss.

Table 1: Average absolute distances (×100) of the proportion of VaR violations to the nominal risk ($\alpha = 0.05$).

	Based on extracted signals					Based on original series			
	SSA	AR	$ARIMA$	ETS	NN	AR	$ARIMA$	ETS	NN
$N(0,1)$	1.473	1.723	1.893	2.033	1.627	1.390	1.823	1.050	2.130
t_4	0.987	1.207	1.313	1.417	1.090	1.100	1.553	0.937	1.540
t_8	1.310	1.523	1.657	1.747	1.443	1.327	1.717	1.010	1.970
Model 1	1.144	1.594	1.817	2.100	1.172	0.844	1.106	0.944	1.456
Model 2a	1.350	1.533	1.567	1.728	1.628	1.689	1.911	1.050	2.394
Model 2b	1.233	1.356	1.550	1.494	1.328	1.372	2.239	0.911	2.067
Model 3	1.611	1.978	1.983	2.094	1.822	1.617	1.989	1.589	1.900
Model 4	0.944	0.961	1.189	1.244	0.983	0.839	1.244	0.500	1.583
Low NSR	1.204	1.391	1.473	1.513	1.338	1.240	1.140	0.904	2.293
High NSR	1.309	1.578	1.769	1.951	1.436	1.304	2.256	1.093	1.467
$h = 1$	1.300	1.633	1.813	1.737	1.533	1.233	1.767	0.957	1.373
$h = 5$	1.107	1.360	1.420	1.557	1.217	1.177	1.587	1.067	1.757
$h = 10$	1.363	1.460	1.630	1.903	1.410	1.407	1.740	0.973	2.510
Total	1.256	1.484	1.621	1.732	1.387	1.272	1.698	0.998	1.880

Note: Averages of 100 times the absolute value of the distance of the frequency of violations to 0.05.

Table 2: Average absolute distances (×100) of the proportion of VaR violations to nominal risk ($\alpha = 0.01$).

	Based on extracted signals					Based on original series			
	SSA	AR	$ARIMA$	ETS	NN	AR	$ARIMA$	ETS	NN
$N(0,1)$	0.980	1.187	1.340	1.433	1.020	0.847	1.533	0.663	1.180
t_4	0.630	0.743	0.850	0.950	0.683	0.550	1.180	0.477	0.787
t_8	0.743	0.970	1.123	1.247	0.913	0.777	1.250	0.583	1.137
Model 1	0.739	1.000	1.083	1.350	0.789	0.528	0.683	0.578	0.761
Model 2a	0.806	1.022	1.211	1.294	1.022	0.867	1.717	0.617	1.278
Model 2b	0.733	0.928	1.122	1.194	0.922	0.883	1.761	0.594	1.289
Model 3	0.978	1.100	1.244	1.411	0.900	0.728	1.261	0.744	0.883
Model 4	0.667	0.783	0.861	0.800	0.728	0.617	1.183	0.339	0.961
Low NSR	0.778	0.898	1.016	1.087	0.891	0.767	0.773	0.553	1.278
High NSR	0.791	1.036	1.193	1.333	0.853	0.672	1.869	0.596	0.791
$h = 1$	0.830	1.207	1.260	1.223	1.013	0.720	1.207	0.593	0.810
$h = 5$	0.720	0.810	0.943	1.077	0.673	0.650	1.310	0.497	0.940
$h = 10$	0.803	0.883	1.110	1.330	0.930	0.803	1.447	0.633	1.353
Total	0.784	0.967	1.104	1.210	0.872	0.719	1.321	0.574	1.034

Note: Averages of 100 times the absolute value of the distance of the frequency of violations to 0.01.

Table 3: Average loss of VaR violations ($\alpha = 0.05$).

	Based on extracted signals					Based on original series			
	SSA	AR	$ARIMA$	ETS	NN	AR	$ARIMA$	ETS	NN
$N(0,1)$	0.0274	0.0249	0.0257	0.0244	0.0260	0.0303	0.0227	0.0226	0.0369
t_4	0.1165	0.1159	0.1154	0.1171	0.1196	0.1318	0.1153	0.1114	0.1367
t_8	0.0454	0.0456	0.0448	0.0471	0.0483	0.0540	0.0441	0.0412	0.0612
Model 1	0.0349	0.0319	0.0334	0.0343	0.0335	0.0359	0.0340	0.0350	0.0405
Model 2a	0.0898	0.0912	0.0911	0.0928	0.0939	0.0947	0.0927	0.0872	0.1069
Model 2b	0.1030	0.1025	0.0999	0.1002	0.1100	0.1401	0.1057	0.0974	0.1435
Model 3	0.0397	0.0399	0.0399	0.0415	0.0410	0.0424	0.0397	0.0405	0.0448
Model 4	0.0482	0.0450	0.0455	0.0454	0.0447	0.0469	0.0318	0.0318	0.0557
Low NSR	0.0884	0.0866	0.0860	0.0875	0.0908	0.1053	0.0811	0.0796	0.1174
High NSR	0.0378	0.0377	0.0379	0.0382	0.0385	0.0387	0.0404	0.0372	0.0391
$h = 1$	0.0589	0.0583	0.0584	0.0587	0.0579	0.0590	0.0509	0.0510	0.0601
$h = 5$	0.0445	0.0446	0.0451	0.0463	0.0474	0.0571	0.0468	0.0449	0.0644
$h = 10$	0.0860	0.0835	0.0824	0.0836	0.0886	0.1000	0.0847	0.0793	0.1103
Total	0.0631	0.0621	0.0619	0.0628	0.0646	0.0720	0.0607	0.0584	0.0782

Note: Quadratic loss of violations.

Table 4: Average loss of VaR violations ($\alpha = 0.01$).

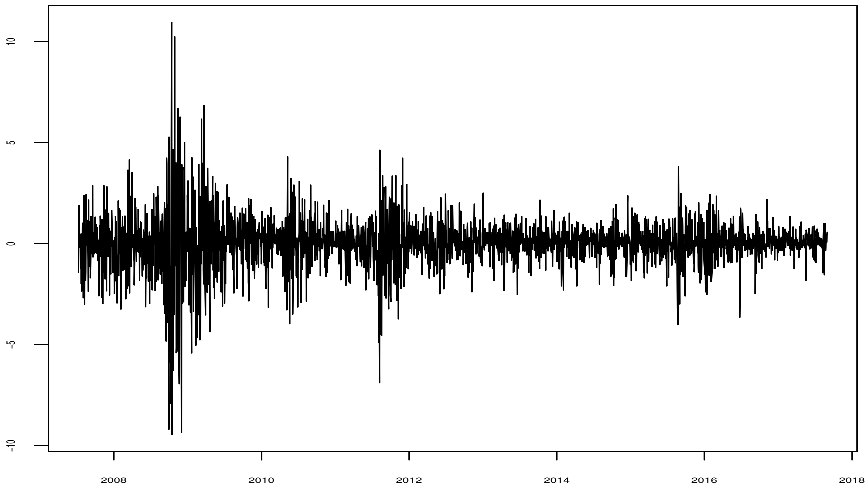
	Based on extracted signals					Based on original series			
	SSA	AR	$ARIMA$	ETS	NN	AR	$ARIMA$	ETS	NN
$N(0,1)$	0.0105	0.0107	0.0125	0.0120	0.0103	0.0113	0.0149	0.0072	0.0161
t_4	0.0618	0.0619	0.0622	0.0635	0.0635	0.0726	0.0636	0.0571	0.0734
t_8	0.0203	0.0212	0.0220	0.0236	0.0221	0.0207	0.0211	0.0157	0.0254
Model 1	0.0174	0.0172	0.0184	0.0200	0.0170	0.0174	0.0157	0.0158	0.0200
Model 2a	0.0539	0.0558	0.0577	0.0592	0.0562	0.0543	0.0625	0.0516	0.0632
Model 2b	0.0458	0.0445	0.0452	0.0452	0.0490	0.0640	0.0540	0.0364	0.0643
Model 3	0.0202	0.0222	0.0225	0.0238	0.0226	0.0213	0.0187	0.0198	0.0216
Model 4	0.0170	0.0166	0.0174	0.0170	0.0150	0.0174	0.0151	0.0096	0.0223
Low NSR	0.0405	0.0399	0.0406	0.0414	0.0419	0.0481	0.0367	0.0336	0.0542
High NSR	0.0212	0.0226	0.0239	0.0247	0.0220	0.0217	0.0297	0.0197	0.0224
$h = 1$	0.0278	0.0300	0.0308	0.0307	0.0287	0.0252	0.0265	0.0222	0.0256
$h = 5$	0.0192	0.0205	0.0211	0.0224	0.0216	0.0254	0.0249	0.0187	0.0297
$h = 10$	0.0456	0.0432	0.0448	0.0460	0.0455	0.0541	0.0482	0.0390	0.0595
Total	0.0308	0.0312	0.0322	0.0330	0.0319	0.0349	0.0332	0.0266	0.0383

Note: Quadratic loss of violations.

6 VaR in SP500

A daily series of returns of the SP500 index from the 10th of July of 2007 to the 31st of August of 2017 (2558 observations) is shown in Figure 1. This interval includes important economic turbulences, such as the subprime crisis, and periods of greater stability. The possible absence of autocorrelation in the returns as imposed in the models in equation (1) is tested with the corrected version of the Box-Pierce statistic as suggested by Deo (2000) and Lobato et al. (2001), which is robust to the presence of higher order dependence typical of financial time series. The corrected Box-Pierce statistic for the first 10, 50 and 100 autocorrelations takes values of 13.31, 47.05 and 99.24 with p-values 0.21, 0.59 and 0.50, confirming the absence of linear correlation in the returns for the usual levels of significance and their agreement with the typical form of the Efficient Market Hypothesis.

Figure 1: SP500 daily returns



The VaR is estimated in this series in 500 rolling samples with $n = 2048$ observations and horizon of predictions $h = 1, 5, 10$ in such a way that the 500 predictions correspond to the more recent observations, regardless the horizon of prediction. The sample size is the same as the one used in the Monte Carlo in the previous section, which allows us to use the fast Fourier transform to calculate the periodogram, implying a significant saving in computational time. The SSA is implemented with the same window lengths $L = 1008$

and $k = 5$ as in the Monte Carlo.

Table 5 shows the proportions of VaR violations, that is the proportion of times when $z_{n+h} < VaR_{n+h|n}(\alpha)$, for $\alpha = 0.05, 0.01$ over the 500 hundred rolling samples. Table 5 also shows the average VaR over these 500 rolling series. The closer this value is to zero the tighter the control of losses for a given confidence, that is the lower the amount of capital required to cover for potential losses for a given confidence. Finally, the magnitude of the effective losses when a violation occurs is also included. An adequate VaR should not only control for the frequency of violations but it is also important for the loss not to be very large when a violation occur. Table 5 shows the average loss (over the 500 rolling samples) calculated as 100 times the quadratic loss defined in (5). For comparative purposes, in addition to the quantile of the Studentised returns, the naive quantile from a standard normal distribution is also considered for $q_\varepsilon(\alpha)$.

Table 5: Comparison of different VaR in SP500

	$\alpha = 0.05$								
	SSA			ETS (series)			GJR-GARCH		
	$h=1$	$h=5$	$h=10$	$h=1$	$h=5$	$h=10$	$h=1$	$h=5$	$h=10$
Viol.	0.038	0.038	0.042	0.044	0.044	0.038	0.034	0.032	0.030
Viol.(S)	0.038	0.038	0.042	0.044	0.046	0.040	0.022	0.030	0.026
Av. VaR	-1.428	-1.440	-1.448	-1.191	-1.207	-1.222	-1.315	-1.402	-1.487
Av. VaR (S)	-1.427	-1.439	-1.447	-1.191	-1.206	-1.222	-1.412	-1.505	-1.597
Av. loss	2.257	2.246	2.468	3.004	3.230	3.418	2.450	2.204	1.969
Av. loss (S)	2.273	2.218	2.455	3.018	3.212	3.413	2.213	1.902	1.660
	$\alpha = 0.01$								
	SSA			ETS (series)			GJR-GARCH		
	$h=1$	$h=5$	$h=10$	$h=1$	$h=5$	$h=10$	$h=1$	$h=5$	$h=10$
Viol.	0.018	0.018	0.020	0.016	0.016	0.018	0.010	0.012	0.012
Viol. (S)	0.014	0.010	0.010	0.014	0.016	0.018	0.008	0.006	0.004
Av. VaR	-2.018	-2.036	-2.047	-1.684	-1.706	-1.728	-1.860	-1.982	-2.103
Av. VaR (S)	-2.356	-2.376	-2.389	-1.966	-1.991	-2.017	-2.205	-2.350	-2.493
Av. loss	0.715	0.670	0.794	1.838	2.051	2.145	1.464	1.050	0.823
Av. loss (S)	0.333	0.303	0.374	1.419	1.568	1.639	1.087	0.696	0.517

Note: h, Viol., Av. VaR and Av. loss denote the prediction horizon, the proportion of violations, the average VaR and the average quadratic loss over the 500 rolling samples using $N(0, 1)$ quantiles. Viol. (S), Av. VaR (S) and Av. loss (S) denote the proportion of violations, average VaRs and average quadratic loss using sample quantiles of the Studentised series.

Based on the performance of the different techniques previously analysed, Table 5 only shows the results obtained with ETS in the observables and SSA. In addition to these two techniques, the VaR is also calculated with the GJR-GARCH(1,1) model proposed by Glosten et al. (1993) in order to analyse the applicability of these simpler techniques (EGARCH(1,1) and GARCH(1,1) models were also considered, giving similar or slightly worse results). The GJR-GARCH model has been selected based on Bams et al. (2017), who show that it performs very well in SP500 daily series and other indexes, and in particular gives better VaR forecasts than other implied volatility and historical volatility models. It is then interesting to see if the VaR estimation techniques proposed in this paper can offer better estimation than the GJR-GARCH, which can then be considered as the benchmark. Note that in this case the Studentisation of the returns used to obtain $q_\varepsilon(\alpha)$ is done with conditional standard deviations obtained in the estimation of the GJR-GARCH model. In the rest of cases SSA is employed.

Table 5 shows that the three strategies overstate the risk at 5%, with a lower number of violations than desired, especially the GJR-GARCH, where the proportion of violations is clearly inferior to 0.05. However, all the techniques tend to understate the risk at 1%, except the GJR-GARCH using the sample quantiles of the Studentised series. Regarding the average VaRs, lower frequencies of violations generally come with larger VaR and lower losses. However, it is noteworthy that SSA with the quantile based on Studentised series may lead to lower losses with larger frequencies of violation closer to the nominal α, especially for $\alpha = 0.01$.

Different backtesting techniques can be used to asses the adequacy of the VaR estimates. Some of the most common tests for backtesting VaRs are:

- The Unconditional Coverage test proposed by Kupiec (1995). It compares the frequency of violations ($\hat{\alpha}$) with the unconditional coverage α with a likelihood ratio test. The test statistic is $2[\log(\hat{\alpha}^x(1-\hat{\alpha})^{500-x} - \log(\alpha^x(1-\alpha)^{500-x})]$ where x is the number of violations. Under the equality of $\hat{\alpha}$ and α the asymptotic distribution of the test statistic is χ^2_1.

- The independence test proposed by Christoffersen (1998). Under a correct specification of the VaR, violations should occur independently of the rest of violations. The test examines whether the probability of a violation depends on the previous outcome. The statistic is $2[\log((1-\pi_0)^{n_{00}}\pi_0^{n_{01}}(1-\pi_1)^{n_{10}}\pi_1^{n_{11}})-\log((1-\pi)^{n_{00}+n_{10}}\pi^{n_{01}+n_{11}})]$ where n_{ij} is the number of times that a j follows a i, where $j, i = 0$ means no violations and $j, i = 1$ denotes violation, $\pi_0 = n_{01}/(n_{00} + n_{01})$, $\pi_1 = n_{11}/(n_{10} + n_{11})$ and $\pi = (n_{01} + n_{11})/(n_{00} + n_{01} + n_{10} + n_{11})$. Under the null hypothesis of independence of violations π_0 and π_1 should be equal and the test statistic has a χ^2_1 limiting distribution.

- Traffic lights (Basel regulations). The Basel Committee on Banking Supervision proposed a Traffic Light approach to analyse the VaR violations in their 1996 document Basel Committee on Banking Supervision (1996). Therein the Basel Committee defines three color zones through cumulative probabilities of the number of realized VaR violations. The Green Zone is defined as the number of violations whereby the cumulative probability of obtaining that many violations or fewer is less than 95% if the VaR is correct. The Yellow Zone is defined as the number of violations whereby the cumulative probability of obtaining that many violations or fewer is greater than 95% but less than 99.99%. Finally, the Red Zone is defined by a cumulative probability of 99.99% or more. Table 6 shows the different zones for $\alpha = 0.05, 0.01$ and 500 observations.

Table 6: Traffic light approach (Basel Committee 1996)

	$\alpha = 0.01$				$\alpha = 0.05$			
Zone	Green	Yellow		Red	Green	Yellow		Red
N. of violations	≤ 8	9	14	≥ 15	≤ 32	33	44	≥ 45
Cumulative prob.	93.29	96.89	99.98	99.99	93.36	95.46	99.97	99.99

Note: Cumulative probability is the probability of obtaining a number violations less or equal to the especified value when the VaR is correct (i.e. true coverage is 100*(1-α)%). Based on a sample of 500 observations. The yellow zone begins where cumulative probability exceeds 95%, and the red zone begins at a cumulative probability of 99.99%.

- Out-of-sample dynamic quantile test. Engle & Manganelli (2004) proposed a test to check the independence of the series of violations of past violations, VaR estimates and other past variables. For a given confidence α, the test statistic is

$(Viol - \alpha)'X(X'X)^{-1}X'(Viol - \alpha)/(\alpha(1-\alpha)$ where $Viol$ is a vector with typical element $Viol_t = 1$ if a violation occurs and 0 otherwise and X is a matrix with columns containing the observations of variables whose orthogonality with $Viol$ is to be tested (e.g. constant, lagged $Viol$, VaR...). The limiting null distribution is χ^2_q, where $q = rank(X)$.

Table 7 shows the results of these backtesting tests in the 500 rolling sub-series of SP500 daily returns. The dynamic quantile test has been obtained with constant term, four lags and the VaR used as instruments, and thus the null limit distribution is χ^2_6. Considering first a confidence of 95%, it is noteworthy that the unconditional coverage test rejects the VaR estimated with GJR-GARCH, which taking into account the widely documented low power of this test suggests against the use of this model, at least for $\alpha = 0.05$. The rest of techniques perform quite well, especially if the Studentised returns are used for $q_\varepsilon(\alpha)$, and the only rejection is for the dynamic quantile test for SSA with $h = 1$. For a confidence of 99%, SSA with Studentised returns does remarkably well, better than the ETS applied on the original series.

Table 7: Backtesting the VaR in SP500

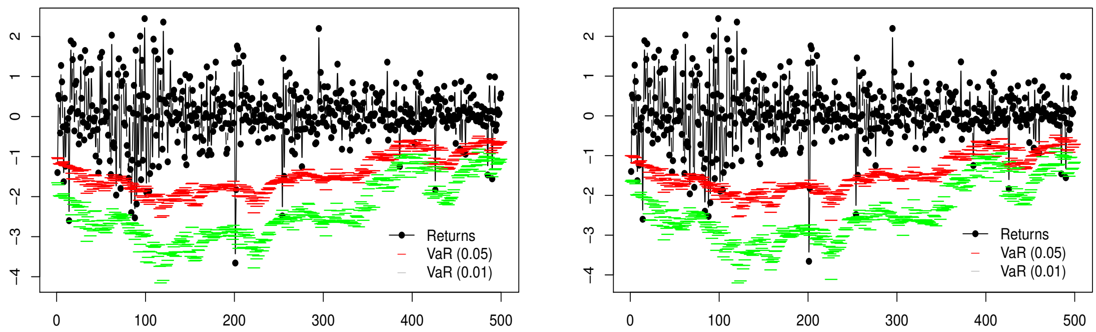
$\alpha = 0.05$									
	SSA			ETS (series)			GJR-GARCH		
	$h=1$	$h=5$	$h=10$	$h=1$	$h=5$	$h=10$	$h=1$	$h=5$	$h=10$
Un. Cov.	1.647	1.647	0.711	0.394	0.394	1.647	3.021*	3.889**	4.884**
Un. Cov.(S)	1.647	1.647	0.711	0.394	0.173	1.127	10.347***	4.884**	7.298***
Ind.	0.138	0.138	0.031	0.001	0.001	1.700	1.120	2.768*	0.537
Ind. (S)	0.138	0.138	0.031	0.001	0.004	1.416	0.496	0.537	0.914
Zone	Green	Green	Green	Green	Green	Green	Green	Green	Green
Zone (S)	Green	Green	Green	Green	Green	Green	Green	Green	Green
DQ	13.278**	6.943	7.321	4.181	5.347	13.481**	3.984	7.651	8.613
DQ (S)	13.079**	6.592	7.044	4.140	5.341	11.153*	8.585	5.826	12.124*
$\alpha = 0.01$									
	SSA			ETS (series)			GJR-GARCH		
	$h=1$	$h=5$	$h=10$	$h=1$	$h=5$	$h=10$	$h=1$	$h=5$	$h=10$
Un. Cov.	2.613	2.613	3.914**	1.538	1.538	2.612*	0.000	0.190	0.190
Un. Cov. (S)	0.719	0.000	0.000	0.719	1.538	2.612*	0.217	0.943	2.353
Ind.	0.331	0.331	0.368	0.261	2.566	2.126	0.101	3.711**	0.146
Ind. (S)	0.199	0.101	0.101	0.199	2.566	2.126	0.065	0.036	0.016
Zone	Yellow	Yellow	Yellow	Green	Yellow	Yellow	Green	Green	Green
Zone (S)	Green	Green	Green	Green	Green	Yellow	Green	Green	Green
DQ	22.171***	16.344**	16.769**	4.181	27.117***	26.819***	3.984	29.699***	1.871
DQ (S)	10.369	6.555	5.711	4.140	26.996***	26.746***	8.585	1.814	2.057

Note: Un. Cov., Ind., Zone and DQ denote respectively the unconditional coverage test (Kupiec, 1995), independence test (Chistoffersen, 1998), the traffic light (Basel regulation, 1996) and the out-of-sample dynamic quantile test (Engle and Manganelli, 2004) with 4 lags of the violations and the estimated VaR as instruments. 500 rolling samples have been used. The symbol (S) denotes that the sample quantiles of the Studentised series have been used to estimate the VaR, whereas no symbol implies that $N(0,1)$ quantiles have been used. *, **, *** denote rejection of the corresponding null at 10%, 5% and 1% respectively.

Finally, Figure 2 shows the 500 SP500 returns used for forecasting evaluation together with the VaR at 95% and 99% confidence levels predicted by SSA for horizons of prediction $h = 1$ and $h = 5$, with the quantiles obtained with the Studentised returns. The figure corroborates the good performance of SSA, showing that the estimated VaR rapidly reacts to the changes in volatility even for a far horizon of prediction $h = 5$.

Figure 2: VaR for SP500

(a) $h = 1$ (b) $h = 5$



7 Conclusion

Estimation of VaR in SV models requires prediction of the conditional variance, which is not a simple task due to the latent character of the factor that drives the volatility. A combination of in-sample prediction based on SSA with other out-of-sample forecasting techniques can be used to estimate the future volatility and the VaR. The Monte Carlo analysis and an application to a real series of daily SP500 returns show that SSA for in-sample and out-of-sample prediction and a combination of SSA for in-sample and ETS for out-of-sample prediction are good options, performing better than GJR-GARCH models, which was shown in Bams et al. (2017) to provide very satisfactory results.

Data availability statement

The empirical data used in this paper can be freely downloaded from different sources. For example from the FRED Economic Data, website https://fred.stlouisfed.org/.

References

Armstrong, J.S. 1985. *Long-range Forecasting: From Crystal Ball to Computer*. 2nd. ed. Wiley. ISBN 978-0-471-82260-8.

Arteche, J., 2004. Gaussian Semiparametric Estimation in Long Memory in Stochastic Volatility and Signal Plus Noise Models. *Journal of Econometrics* 119, 131-154.

Arteche, J., 2015. Signal extraction in Long Memory Stochastic Volatility. *Econometric Theory* 31, 1382-1402.

Arteche, J. and García-Enríquez, J. 2017. Singular Spectrum Analysis for signal extraction in Stochastic Volatility models. *Econometrics and Statistics* 1, 85-98.

Dennis Bams, D., Blanchard, G. and Lehnert, T. (2017) Volatility measures and Value-at-Risk. *International Journal of Forecasting*, 33, 848-863.

Barone-Adesi, G., Bourgoin, F., Giannopoulos, K., 1998. Don't look back. *Risk* 11.

Basle Committee on Banking Supervision. 1996. Supervisory Framework For The Use of Back-Testing in Conjunction With The Internal Models Approach to Market Risk Capital Requirements. Available online: www.bis.org/publ/bcbs22.htm

Beneki, C. Eeckels, B. and Leon, C., 2012. Signal extraction and forecasting of the UK Tourism Income time series: A Singular Spectrum Analysis approach. *Journal of Forecasting* 31, 391-400.

Bógalo J., Poncela P. and Senra E., 2021. Circulant singular spectrum analysis: A new automated procedure for signal extraction. *Signal Processing* 179. DOI: 10.1016/j.sigpro.2020.107824.

Bollerslev, T. and H.O Mikkelsen, 1996. Modeling and Pricing Long Memory in Stock Market Volatility. *Journal of Econometrics* 73, 151-184.

Breidt, F.J., Crato, N. and P. de Lima, 1998. The Detection and Estimation of Long Memory in Stochastic Volatility. *Journal of Econometrics* 83, 325-348.

Broomhead, D. and G. King, 1986. Extracting qualitative dynamics from experimental data. *Physica D* 20, 217-236.

de Carvalho, M., Rodrigues, P.C. and Rua, A., 2012. Tracking the US business cycle with a singular spectrum analysis. *Economics Letters* 114, 32-35.

Christofferssen, P. 1998. Evaluating Interval Forecasts. *International Economic Review* 39, 841-862.

Deo, R.S., 2000. Spectral tests of the martingale hypothesis under conditional heteroscedasticity. *Journal of Econometrics* 99, 291-315.

Deo, R.S. and C.M. Hurvich, 2001. On the log periodogram regression estimator of the memory parameter in long memory stochastic volatility models. *Econometric Theory* 17, 686-710.

Engle, R., and Manganelli, S. 2004. CAViaR: Conditional Autoregressive Value-at-Risk by Regression Quantiles. *Journal of Business and Economic Statistics* 22, 367-381.

Glosten, L. R., Jagannathan, R. and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5), 1779-1801.

Golyandina N. E. 2010. On the choice of parameters in Singular Spectrum Analysis and related subspace-based methods. *Statistics and Interface* 3, 259-279.

Golyandina N. E., Nekrutkin V. V. and A. A. Zhigljavsky, 2001. *Analysis of Time Series Structure: SSA and Related Techniques*. Boca Raton, FL: Chapman & Hall/CRC.

González-Rivera, G., Lee, T.H. and Mishra, S. (2004) Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. *International Journal of Forecasting* 20, 629-645.

Granger, C.W.J. and N. Hyung, 1996. Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance* 11, 399–421.

Hassani, H., Heravi, S. and Zhigljavsky, A., 2009. Forecasting European industrial production with Singular Spectrum Analysis. *International Journal of Forecasting* 25, 103-118.

Hassani, H., Heravi, S. and Zhigljavsky, A., 2013. Forecasting UK Industrial Production with Multivariate Singular Spectrum Analysis. *Journal of Forecasting* 32, 395-408.

Harvey, A.C., 1998. Long memory in stochastic volatitility. In: Knight, J., Satchell, S. (Eds.), *Forecasting Volatility in Financial Markets*, Oxford: Butterworth-Haineman, 307-320.

Harvey, A.C., Ruiz, E. and N. Shephard, 1994. Multivariate Stochastic Variance Models. *Review of Economic Studies* 61, 247-264.

Hassani,H., Heravic, S. and A. Zhigljavskya, 2009. Forecasting European industrial production with singular spectrum analysis. *International Journal of Forecasting* 25, 103–118.

Hurvich, C.M., Moulines, E. and P. Soulier, 2005. Estimating long memory in volatility. *Econometrica* 73, 1283-1328.

Hyndman, R.J. and Koehler, A.B., 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting* 22, 679-688.

Hyndman, R.J. and Athanasopoulos, G. 2014. *Forecasting: principles and practice.* OTexts.

Hyndman, R.J. and Khandakar, Y. 2008. Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, 27, 1-22.

Hyndman, R.J., Koehler, A.B., Snyder, R.D., and Grose, S. 2002. A state space framework for automatic forecasting using exponential smoothing methods. *International J. Forecasting*, 18, 439-454.

Khan, M.A.R. and D.S. Poskitt, 2017. Forecasting stochastic processes using singular spectrum analysis: Aspects of the theory and application. *International Journal of Forecasting* 33, 199-213.

Kupiec, P. 1995. Techniques for Verifying the Accuracy of Risk Management Models. *Journal of Derivatives* 3, 73-84.

Liesenfeld, R. and Jung, R.C. 2000. Stochastic Volatility models: Conditional normality versus heavy-tailed distributions. *Journal od Applied Econometrics* 15, 137-160.

Lisi, F. and Medio, A. (1997). Is a random walk the best exchange rate predictor?. *International Journal of Forecasting* 13, 255-267.

Lobato, I., Nankervis, J.C. and N.E. Savin, 2001. Testing for autocorrelation using a modified Box-Pierce Q test. *International Economic Review* 42, 187-205.

Lobato, I.N. and N.E. Savin, 1998. Real and spurious long-memory properties of stock-market data. *Journal of Business and Economic Statistics* 16, 261– 268.

McNeil, A.J. and Frey, R.(2000) Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance* 7, 271-300.

Papailias, F. and Thomakos, D., 2017. EXSSA: SSA-based reconstruction of time series via exponential smoothing of covariance eigenvalues. *International Journal of Forecasting* 33, 214-229.

Perron, P. and Z. Qu, 2010. Long-Memory and Level Shifts in the Volatility of Stock Market Return Indices. *Journal of Business and Economic Statistics* 28, 275–290.

Qu, Z. and P. Perron, 2013. A stochastic volatility model with random level shifts and its application to S&P500 and NASDAQ return indices. *Econometrics Journal* 16, 309-339.

Silva, E.S. and Hassani, H., 2015. On the use of Singular Spectrum Analysis for forecasting U.S. trade before, during and after the 2008 recession. *International Economics* 141, 34-49.

Soofi A.S., Cao L., 2002. Nonlinear Forecasting of Noisy Financial Data. In: Soofi A.S., Cao L. (eds) *Modelling and Forecasting Financial Data. Studies in Computational Finance*, vol 2. Springer, Boston, MA, 455-465. DOI:10.1007/978-1-4615-0931-8-22.

Taylor, S.J., 1986. *Modelling Financial Time Series*, Chichester: John Wiley and Sons.

Vautard, R. and Ghil, M., 1989. Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Physica D* 35, 395-424.

Vautard, R., Yiou, P. and Ghil, M., 1992. Singular-spectrum analysis: A toolkit for short, noisy chaotic signal. *Physica D*, 58, 95-126.