Econometrics and Statistics 1 (2017) 85–98

Contents lists available at ScienceDirect

Econometrics and Statistics

journal homepage: www.elsevier.com/locate/ecosta

Singular Spectrum Analysis for signal extraction in Stochastic Volatility models☆

Josu Arteche*, Javier García-Enríquez

Department of Econometrics and Statistics, University of the Basque Country UPV/EHU, Bilbao, Spain

ARTICLE INFO

Article history:
Received 29 December 2015
Revised 6 September 2016
Accepted 6 September 2016
Available online 8 November 2016

Keywords:
Stochastic Volatility
Singular Spectrum Analysis

ABSTRACT

Estimating the in-sample volatility is one of the main difficulties that face Stochastic Volatility models when applied to financial time series. A non-parametric strategy based on Singular Spectrum Analysis is proposed to solve this problem. Its main advantage is its generality as it does not impose any parametric restriction on the volatility component and only some spectral structure is needed to identify it separately from noisy components. Its convincing performance is shown in an extensive Monte Carlo analysis that includes stationary and nonstationary long memory, short memory and level shifts in the volatility component, which are models often used for financial time series. Its applicability is finally illustrated in a daily Dow Jones Industrial index series and an intraday series from the Spanish Ibex35 stock index.

1. Introduction

Modelling volatility, usually characterised by second order moments, has sparked great interest from just before the turn of the 21st century and has been one of the most active fields of research in financial literature. Two main approaches have been used for that purpose: AutoRegressive Conditional Heteroscedasticity (ARCH) models and subsequent extensions, which are characterised by a conditional variance that is fully driven by past observations; and Stochastic Volatility (SV) models, where the volatility is a latent stochastic process driven by innovations that are inherent to the volatility process. This difference has two effects. First, since innovations enter SV models non-linearly, it is not possible to obtain an analytical expression of the likelihood, making it more complicated to estimate SV models. Second, the series of volatilities is far harder to extract in SV models than in ARCH extensions. Conditional variances in ARCH-based models are exact functions of past observations and can be exactly predicted (at least within the sample) as long as the relevant parameters are known (and in any case they can easily be estimated). The latent nature of the volatility in SV models renders its extraction far more difficult and signal extraction techniques are usually required.

SV models are non-linear, but can be linearised by taking logarithms of the squares, after which they take the form of a sum of two components: the component leading the volatility, which is the component of interest or signal, and an added white noise. The volatility component can then be estimated by applying a filter for signal extraction. Parametric filters, such as the Kalman filter (Harvey et al., 1994; Ruiz, 1994) or the optimal Wiener–Kolmogorov filter (Harvey, 1998)

☆ Research supported by Spanish Ministry of Economy and Competitiveness grant ECO2013-40935-P and Econometrics Research Group IT642-13.

* Corresponding author.
E-mail address: josu.arteche@ehu.es (J. Arteche).

http://dx.doi.org/10.1016/j.ecosta.2016.09.004

can be used for that purpose, but their performance and reliability depend on parametric restrictions and misspecification could possibly render the estimation of the signal incorrect. Semiparametric filters, such as the local Wiener–Kolmogorov in Arteche (2015a, 2015b) relax the need for a complete, correct specification of the model but they still need to impose a partial behaviour in the series that cannot be ignored.

This paper instead proposes a fully non-parametric, model-free, technique for signal extraction based on Singular Spectrum Analysis (SSA). Its main advantages are its flexibility and generality due to its nonparametric nature and the fact that there is no need to impose any statistical restriction on the components to be extracted, which is a clear advantage over other filters for signal extraction such as those mentioned above.

SSA is a relatively novel nonparametric technique for times series analysis. It is based on decomposing a time series as the sum of a finite number of elements (see Golyandina et al., 2001, for a detailed description). After appropriate grouping, these elements can be identified as interpretable components of the time series and the volatility component can then be estimated by selecting those components that share similar characteristics with those in the signal to be estimated. We propose assessing that similarity using spectral tools. Taking into account that the spectral density function of the added noise is flat, all the relevant structure in the spectrum of the log of squares is due to the signal. The elements in the SSA are thus selected as those sharing a similar structure in their spectrum. This strategy has the main advantage of not requiring any parametric restrictions in the volatility component. It therefore allows for different forms of stationary and non-stationary volatilities, covering most of the processes that have been used to model volatility in financial time series. Only some spectral structure is required to identify the signal separately from the added noise characterised by a flat spectrum. Low frequency spectral concentration, caused for example by persistent volatility and/or by level shifts, and seasonal or cyclical spectral peaks are the most frequent distinctive behaviours and those are precisely the situations on which we focus.

The rest of the paper is organised as follows. Section 2 reviews the structure and characteristics of SV models. Section 3 introduces the strategy based on SSA that we propose for signal extraction and volatility estimation. Section 4 analyses its performance in a Monte Carlo analysis. Section 5 shows the applicability of the proposed technique for estimating the volatility of two real financial time series: a daily series of returns from the Dow Jones Industrial index and a series of intraday returns from the Spanish Ibex35 stock index. Finally, Section 6 concludes.

2. Stochastic Volatility models

SV models are defined as

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

where $\sigma_t = \sigma \exp(v_t/2)$ for σ a positive constant scale factor, v_t is the volatility component and $\varepsilon_t \sim iid(0, 1)$ (see Taylor, 1986). Taking the logs of the squares of z_t in (1) we have

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

where $\mu = \log \sigma^2 + E \log \varepsilon_t^2$ and $\xi_t = \log \varepsilon_t^2 - E \log \varepsilon_t^2$ 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 $\sigma_\xi^2 = \pi^2/2$. Apart from the constant μ, y_t takes the form of a signal plus noise and the volatility component v_t can be estimated by means of signal extraction techniques.

Uncorrelation between signal and noise is usually assumed such that the autocovariance function of y_t is

$$\gamma_y(h) = E y_t y_{t+h} = \gamma_v(h) + \sigma_\xi^2 I_{h=0}, \tag{3}$$

where $I_{h=0} = 1$ if $h = 0$ and 0 otherwise. Consequently, the autocovariances of y_t coincide with those of the signal v_t and only the variance is affected by the noise. Note that uncorrelation between v_t and ξ_s does not preclude the existence of leverage in the form of correlation between ε_s and v_t, which may be allowed on a non contemporaneous basis to maintain the martingale difference character of z_t. In fact, as long as the joint distribution of ε_s and v_t 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). Higher order dependencies between v_t and ξ_s are also possible (see Arteche, 2015a). Under these conditions the spectral density function of y_t is

$$f_y(\lambda) = f_v(\lambda) + \frac{\sigma_\xi^2}{2\pi} \qquad \text{for } -\pi \le \lambda \le \pi, \tag{4}$$

and then any structure in $f_y(\lambda)$ in the form of departures from being constant is due to the signal v_t and can be used to identify the volatility component separately from the added noise.

Originally, SV models assumed that v_t was a stationary process, and the AR(1) was the process that attracted most interest in those first attempts (see Harvey et al., 1994, among many others). An estimate of the autoregressive parameter that is positive and close to one is usually obtained when this model is fitted to real financial time series, which implies high persistence characterised by power concentration around frequency zero in $f_v(\lambda)$. This high persistence found in most financial series led many authors to propose Long Memory in Stochastic Volatility (LMSV) models, where the power concentration in